US008099674B2

(12) **United States Patent**
Mackinlay et al.

(10) **Patent No.: US 8,099,674 B2**
(45) **Date of Patent: Jan. 17, 2012**

(54) **COMPUTER SYSTEMS AND METHODS FOR AUTOMATICALLY VIEWING MULTIDIMENSIONAL DATABASES**

(75) Inventors: **Jock Douglas Mackinlay**, Bellevue, WA (US); **Christopher Richard Stolte**, Seattle, WA (US); **Patrick Hanrahan**, Portola Valley, CA (US)

(73) Assignee: **Tableau Software LLC**, Seattle, WA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 450 days.

(21) Appl. No.: **11/223,658**

(22) Filed: **Sep. 9, 2005**

(65) **Prior Publication Data**

US 2007/0061611 A1 Mar. 15, 2007

(51) **Int. Cl.**

G06F 3/048	(2006.01)
G06F 3/00	(2006.01)
G06F 17/22	(2006.01)
G06Q 40/00	(2006.01)

(52) **U.S. Cl.** **715/764**; 715/810; 715/700; 715/215; 705/38

(58) **Field of Classification Search** 715/764, 715/215, 212, 700, 810; 705/38
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,800,810 A		1/1989	Masumoto	99/519
5,036,314 A		7/1991	Barillari et al.	340/717
5,060,980 A		10/1991	Johnson et al.	283/70
5,143,888 A		9/1992	Olbrich	502/329
5,144,452 A		9/1992	Abuyama	358/296
5,169,713 A		12/1992	Kumurdjian	428/323
5,265,244 A		11/1993	Ghosh et al.	395/600
5,265,246 A		11/1993	Li et al.	395/600
5,377,348 A		12/1994	Lau et al.	395/600
5,383,029 A		1/1995	Kojima	358/403
5,461,708 A	*	10/1995	Kahn	345/440
5,560,007 A		9/1996	Thai	395/600
5,577,241 A		11/1996	Spencer	395/605
5,581,677 A	*	12/1996	Myers et al.	345/440
5,664,172 A		9/1997	Antoshenkov	395/604
5,664,182 A		9/1997	Nierenberg et al.	395/613
5,668,987 A		9/1997	Schneider	395/603

(Continued)

FOREIGN PATENT DOCUMENTS

HU 215657 B 1/1994

(Continued)

OTHER PUBLICATIONS

Screen Dumps for Microsoft Office Excel 2003 SP2, figures 1-24, 10 pp. 1-19.*

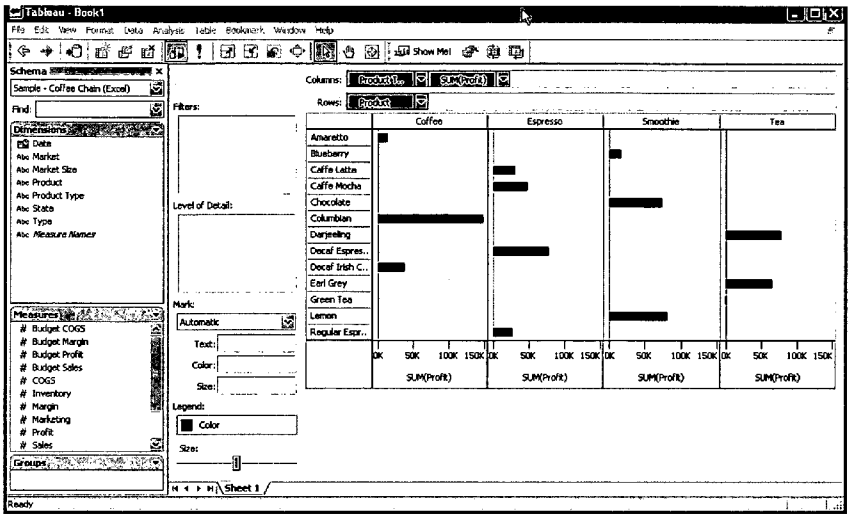
(Continued)

Primary Examiner — Simon Ke
Assistant Examiner — Erik Stitt
(74) *Attorney, Agent, or Firm* — Morgan, Lewis & Bockius LLP

(57) **ABSTRACT**

A method for automatically forming the clearest and most useful visual plot for a given dataset of tuples. A best view type is selected for a view that includes a subsequently added new field. The visual plot is populated with the data in the view and then automatically rendered for the user. A dataset that is retrieved from a storage is analyzed to identify all the data types found in the dataset, and to determine the best view type to assign to the dataset's views. The visual plot is then populated with the data according to this best view type, and is automatically rendered for the user.

202 Claims, 18 Drawing Sheets



U.S. PATENT DOCUMENTS

5,794,246	A	8/1998	Sankaran et al.	707/101
5,864,856	A	1/1999	Young	707/100
5,893,088	A	4/1999	Hendricks et al.	707/3
5,933,830	A	8/1999	Williams	707/100
6,031,632	A	2/2000	Yoshihara et al.	358/403
6,032,158	A	2/2000	Mukhopadhyay et al.	707/201
6,044,374	A	3/2000	Nesamoney et al.	707/10
6,100,901	A	8/2000	Mohda et al.	345/440
6,115,744	A	9/2000	Robins et al.	709/227
6,154,766	A	11/2000	Yost et al.	709/201
6,173,310	B1	1/2001	Yost et al.	709/201
6,188,403	B1	2/2001	Sacerdoti et al.	345/339
6,208,990	B1	3/2001	Suresh et al.	707/6
6,222,540	B1	4/2001	Sacerdoti	345/339
6,247,008	B1	6/2001	Cambot et al.	707/3
6,253,257	B1	6/2001	Dundon	709/331
6,260,050	B1	7/2001	Yost et al.	707/501
6,269,393	B1	7/2001	Yost et al.	709/201
6,300,957	B1	10/2001	Rao et al.	345/441
6,301,579	B1	10/2001	Becker	707/102
6,317,750	B1	11/2001	Tortolani et al.	707/103
6,327,628	B1	12/2001	Anuff et al.	709/311
6,339,775	B1	1/2002	Zamanian et al.	707/101
6,377,259	B1	4/2002	Tenev et al.	345/440
6,397,195	B1	5/2002	Pinard et al.	705/30
6,400,366	B1	6/2002	Davies et al.	345/440
6,405,195	B1	6/2002	Ahlberg	707/4
6,405,208	B1	6/2002	Raghavan et al.	707/102
6,424,933	B1	7/2002	Agrawala et al.	703/2
6,490,593	B2	12/2002	Proctor	707/102
6,492,989	B1	12/2002	Wilkinson	345/440
6,505,205	B1 *	1/2003	Kothuri et al.	1/1
6,522,342	B1	2/2003	Gagnon et al.	345/716
6,529,217	B1 *	3/2003	Maguire et al.	715/769
6,590,577	B1 *	7/2003	Yonts	345/440
6,611,825	B1	8/2003	Billheimer et al.	706/45
6,643,646	B2	11/2003	Su et al.	707/6
6,701,485	B1 *	3/2004	Igra et al.	715/210
6,707,454	B1 *	3/2004	Barg et al.	345/440
6,714,897	B2	3/2004	Whitney et al.	702/189
6,725,230	B2	4/2004	Ruth et al.	707/102
6,750,864	B1	6/2004	Anwar	345/440
6,763,308	B2	7/2004	Chu et al.	702/19
6,768,986	B2	7/2004	Cras et al.	707/2
6,867,788	B1 *	3/2005	Takeda	345/630
6,906,717	B2 *	6/2005	Couckuyt et al.	345/440
6,920,608	B1 *	7/2005	Davis	715/209
6,928,433	B2 *	8/2005	Goodman et al.	1/1
6,985,905	B2 *	1/2006	Prompt et al.	707/829
7,009,609	B2	3/2006	Miyadai	345/440
7,089,266	B2 *	8/2006	Stolte et al.	707/104.1
7,117,058	B2 *	10/2006	Lin et al.	700/108
7,168,035	B1 *	1/2007	Bell et al.	715/234
7,181,450	B2	2/2007	Malloy et al.	707/4
7,250,951	B1 *	7/2007	Hurley et al.	345/440
7,315,305	B2 *	1/2008	Crotty et al.	345/440
7,379,601	B2	5/2008	Yang et al.	382/224
7,630,971	B2 *	12/2009	Arrouye et al.	1/1
7,694,278	B2 *	4/2010	Pasumansky et al.	717/125
7,725,483	B2 *	5/2010	Poyourow et al.	707/763
2002/0016699	A1	2/2002	Hoggart et al.	703/2
2002/0118192	A1	8/2002	Couckuyt et al.	345/440
2002/0123865	A1	9/2002	Whitney et al.	702/189
2002/0154118	A1	10/2002	McCarthy et al.	345/440
2003/0042928	A1	3/2003	Tsai	324/769
2003/0200034	A1	10/2003	Fellenberg et al.	702/20
2003/0204511	A1 *	10/2003	Brundage et al.	707/100
2004/0183800	A1	9/2004	Peterson	345/440
2004/0224577	A1	11/2004	Kaji	440/1
2004/0227759	A1	11/2004	McKnight et al.	345/440
2004/0243593	A1 *	12/2004	Stolte et al.	707/100
2005/0035966	A1	2/2005	Pasquarette et al.	345/440
2005/0035967	A1	2/2005	Joffrain et al.	345/440
2005/0060300	A1 *	3/2005	Stolte et al.	707/3
2005/0099423	A1	5/2005	Brauss	345/440
2005/0234688	A1	10/2005	Pinto et al.	703/6
2005/0234920	A1	10/2005	Rhodes	707/10
2006/0020586	A1 *	1/2006	Prompt et al.	707/3
2006/0031187	A1 *	2/2006	Pyrce et al.	707/1
2006/0053363	A1	3/2006	Bargh et al.	715/503
2006/0100873	A1	5/2006	Bittner et al.	704/256.2
2006/0101391	A1 *	5/2006	Ulke et al.	717/107
2006/0129913	A1	6/2006	Vigesaa et al.	715/503
2006/0136825	A1 *	6/2006	Cory et al.	715/700
2006/0206512	A1 *	9/2006	Hanrahan et al.	707/102
2007/0055487	A1	3/2007	Habitz et al.	703/19
2007/0061344	A1	3/2007	Dickerman et al.	707/100
2007/0112844	A1 *	5/2007	Tribble et al.	707/102
2008/0133573	A1	6/2008	Haft et al.	707/102

FOREIGN PATENT DOCUMENTS

WO WO2006/-060773 * 6/2006

OTHER PUBLICATIONS

Becker et al., "Trellis Graphics Displays: A Multi-Dimensional Data Visualization Tool for Data Mining," Third Annual Conference on Knowledge Discovery in Databases, Aug. 1997.

Becker, "Visualizing Decision Table Classifiers", Proceedings IEEE Symposium on Information Visualization, pp. 102-105, 1998.

Bosch et al., "Performance Analysis and Visualization of Parallel Systems Using SimOS and Rivet: A Case Study," Proceedings of the Sixth IEEE International Symposium on High-Performance Computer Architecture, pp. 360-371, Jan. 2000.

Bosch et al., "Rivet: A Flexible Environment for Computer Systems Visualization," Computer Graphics.34, pp. 68-73, Feb. 2000.

Brunk et al., "MineSet: An Integrated System for Data Mining," Proceeding of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 135-138, AAA1 Press, 1997.

Derthick et al., "An Interactive Visual Query Environment for Exploring Data," Proceedings of ACM SIGGRAPH Symposium on User Interface Software & Technology, pp. 189-198, 1997.

Freeze, Unlocking OLAP with Microsoft SQL Server and Excel 2000, IDG Books Worldwide, Foster City, CA, pp. 155-332 and 379-422.

Fua et al., "Navigating Hierarchies with Structure-Based Brushes," Proc. of Infovis '99 (San Francisco, CA, USA, 1999), IEEE Computer Soc Press.

Goldstein et al., "A Framework for Knowledge-Based Interactive Data Exploration," Journal of Visual Languages and Computing 5, pp. 339-363, Dec. 1994.

Gray et al., "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total," Data Mining and Knowledge Discovery 1, pp. 29-53, 1997.

Healey, "On the Use of Perceptual Cues and Data Mining for Effective Visualization of Scientific Datasets," Proceedings Graphics Interface '98, pp. 177-187.

Kohavi, "Data Mining and Visualization," Frontiers of Engineering: Reports of leading-Edge Engineering from the 200 NAE Sym.

Livny et al., "DEVise: Integrated Querying and Visual Exploration of Large Datasets," Proceedings of ACM SIGMOD, pp. 301-312, May 1997.

MacDonald, "Creating Basic Charts," Excel 2003, Chapter 9, pp. 298-342, 2006.

MacKinlay, J., "Automating the Design of Graphical Presentations of Relational Information," ACM Transactionss on Graphics 5(2), pp. 110-141, 1986.

Perlin et al., "An Alternative Approach to the Computer Interface," Proc. of the 20th International Conference on Computer Graphics and Interactive Techniques, pp. 57-64, 1993.

Rao et al., "The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information," Proc. of ACM SIGCHI, Apr. 1994, pp. 318-322.

Roth et al., "Interactive Graphics Design Using Automatic Presentation Knowledge," Proceedings of the Conference on Human Factors in Computer Systems, Proc. SIGCHI '94, pp. 112-117, 1994.

Roth et al., "Visage: A User Interface Environment for Exploring Information" Proceedings of Information Visualization Symposium 1996, pp. 3-12, Mar. 12, 1999.

Spenke et al., Focus: The Interactive Table for Product Comparison and Selection, Proc. of ACM Symposium on User Interface Software and Technology, Nov. 1996.

Stevens, "On the Theory of Scales of Measurement," Science vol. 103, No. 2684 pp. 677-680, Jun. 7, 1946.
Stolte et al, "Query, Analysis, and Visualization of Hierarchically Structured Data Using Polaris," Proceedings of the Eight ACM SIGKDD '02, International Conference on Knowledge Discovery and Data Mining, Jul. 2002.
Stolte et al., "Multiscale Visualization Using Data Cubes," Proceedings of the Eighth IEEE Symposium on Information Visualization, 2002.
Stolte et al., "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases," IEEE Transactions on Visualization and Computer Graphics 8(1), pp. 52-65, Jan. 2002.
Stolte et al., "Visualizing Application Behavior on Superscalar Processors," Proceedings of the IEEE Symposium on Information Visualization, pp. 10-17, 1999.
Thearling et al., "Visualizing Data Mining Models," Pub. in Information Visualization in Data Mining and Knowledge Discovery, Fayyad, Grinstein and Wierse, eds., Morgan Kaufman, 2001.
Welling et al., "Visualization of Large Multi-Dimensional Datasets," arXiv: astro-ph/0008186, Aug. 11, 2000.
Wilkinson et al., "nViZn: An Algebra-Based Visualization System," Smart Graphics Symposium UK, Mar. 21-23, 2001, Hawthorne, NY USA.
Wilkinson, "Statistics and Computing—The Grammar of Graphics," Springer-Verlag, Inc., New York, 1999.
Notice of Allowance for U.S. Appl. No. 10/453,834 dated Mar. 27, 2006.
Office Action for U.S. Appl. No. 10/667,194 dated Jun. 26, 2006.
Office Action for U.S. Appl. No. 10/667,194 dated Jan. 18, 2007.
Office Action for U.S. Appl. No. 10/667,194 dated Aug. 14, 2007.
Office Action for U.S. Appl. No. 10/667,194 dated Jan. 7, 2008.
Office Action for U.S. Appl. No. 10/667,194 dated Aug. 14, 2008.
Office Action for U.S. Appl. No. 10/667,194 dated Feb. 9, 2009.
Office Action for U.S. Appl. No. 11/005,652 dated Dec. 27, 2007.
Office Action for U.S. Appl. No. 11/005,652 dated Jul. 24, 2008.
Office Action for U.S. Appl. No. 11/005,652 dated Feb. 20, 2009.
Office Action for U.S. Appl. No. 11/488,407 dated Apr. 3, 2009.
Office Action for U.S. Appl. No. 11/787,761 dated Jun. 12, 2008.
Office Action for U.S. Appl. No. 11/787,761 dated Dec. 17, 2008.
International Search Report and Written Opinion for PCT/US2004/30396 dated Aug. 24, 2006.
International Preliminary Report on Patentability for PCT/US2004/30396 dated Apr. 19, 2007.
International Search Report and Written Opinion for PCT/US2005/043937 dated Apr. 18, 2007.
International Preliminary Report on Patentability for PCT/US2005/043937 dated Jun. 14, 2007.
International Search Report and Written Opinion for PCT/US2007/009810 dated Jul. 7, 2008.
International Preliminary Report on Patentability for PCT/US2007/009810 dated Oct. 30, 2008.
International Search Report and Written Opinion for PCT/US2006/35300 dated Jul. 7, 2008.
International Search Report and Written Opinion for PCT/US04/18217 dated Feb. 7, 2006.
International Preliminary Report on Patentability for PCT/US04/18217 dated Oct. 19, 2006.
Hungarian Search Report for HU P0700460 dated Oct. 9, 2007.
Supplementary European Search Report for EP 04754739.3, dated Dec. 17, 2007.
Specification for U.S. Appl. No. 10/453,834, filed Jun. 2, 2003.
Specification for U.S. Appl. No. 11/005,652, filed Dec. 2, 2004.
Specification for U.S. Appl. No. 11/223,658, filed Sep. 5, 2005.
Notice of Allowance, U.S. Appl. No. 11/488,407, Dec. 29, 2009, 8 pages.
Notice of Allowance, U.S. Appl. No. 11/005,652, May 18, 2010, 4 pages.
Notice of Allowance, U.S. Appl. No. 10/667,194, Mar. 5, 2010, 4 pages.
Office Action, U.S. Appl. No. 10/667,194, Sep. 25, 2009, 10 pages.
Office Action, U.S. Appl. No. 11/005,652, Oct. 23, 2009, 14 pages.
Office Action, U.S. Appl. No. 11/787,761, Jul. 8, 2009, 14 pages.
Office Action, U.S. Appl. No. 11/787,761, Nov. 30, 2009, 14 pages.
Office Action, U.S. Appl. No. 11/787,761, Apr. 28, 2010, 10 pages.
Office Action, U.S. Appl. No. 11/787,761, Sep. 14, 2010, 12 pages.
International Preliminary Report on Patentability, PCT/US2006/035300, Mar. 24, 2009, 5 pages.

* cited by examiner

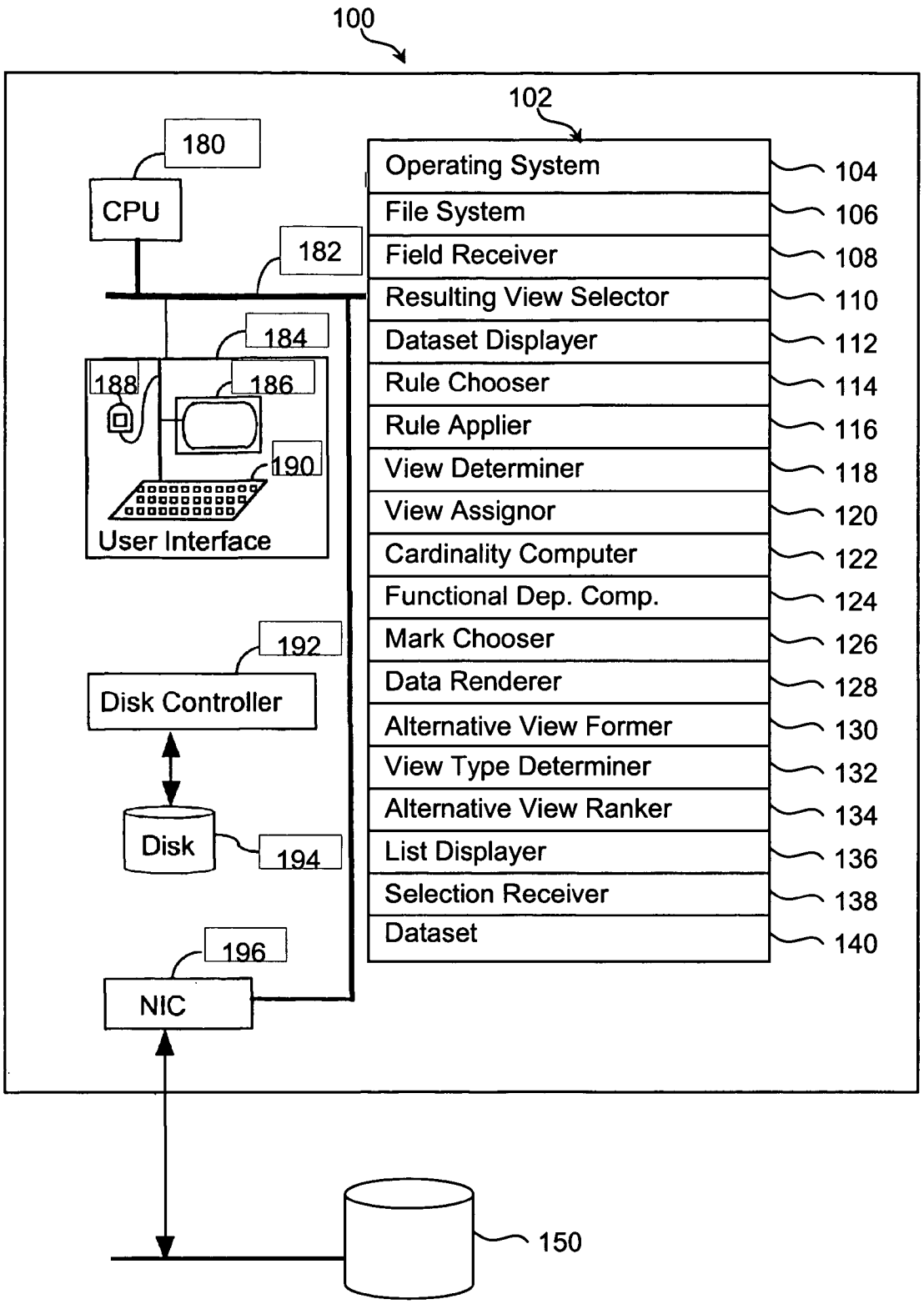

100
102
180
CPU
182
184
188
186
190
User Interface
192
Disk Controller
Disk
194
196
NIC
150
Operating System
File System
Field Receiver
Resulting View Selector
Dataset Displayer
Rule Chooser
Rule Applier
View Determiner
View Assignor
Cardinality Computer
Functional Dep. Comp.
Mark Chooser
Data Renderer
Alternative View Former
View Type Determiner
Alternative View Ranker
List Displayer
Selection Receiver
Dataset
104
106
108
110
112
114
116
118
120
122
124
126
128
130
132
134
136
138
140

Figure 1

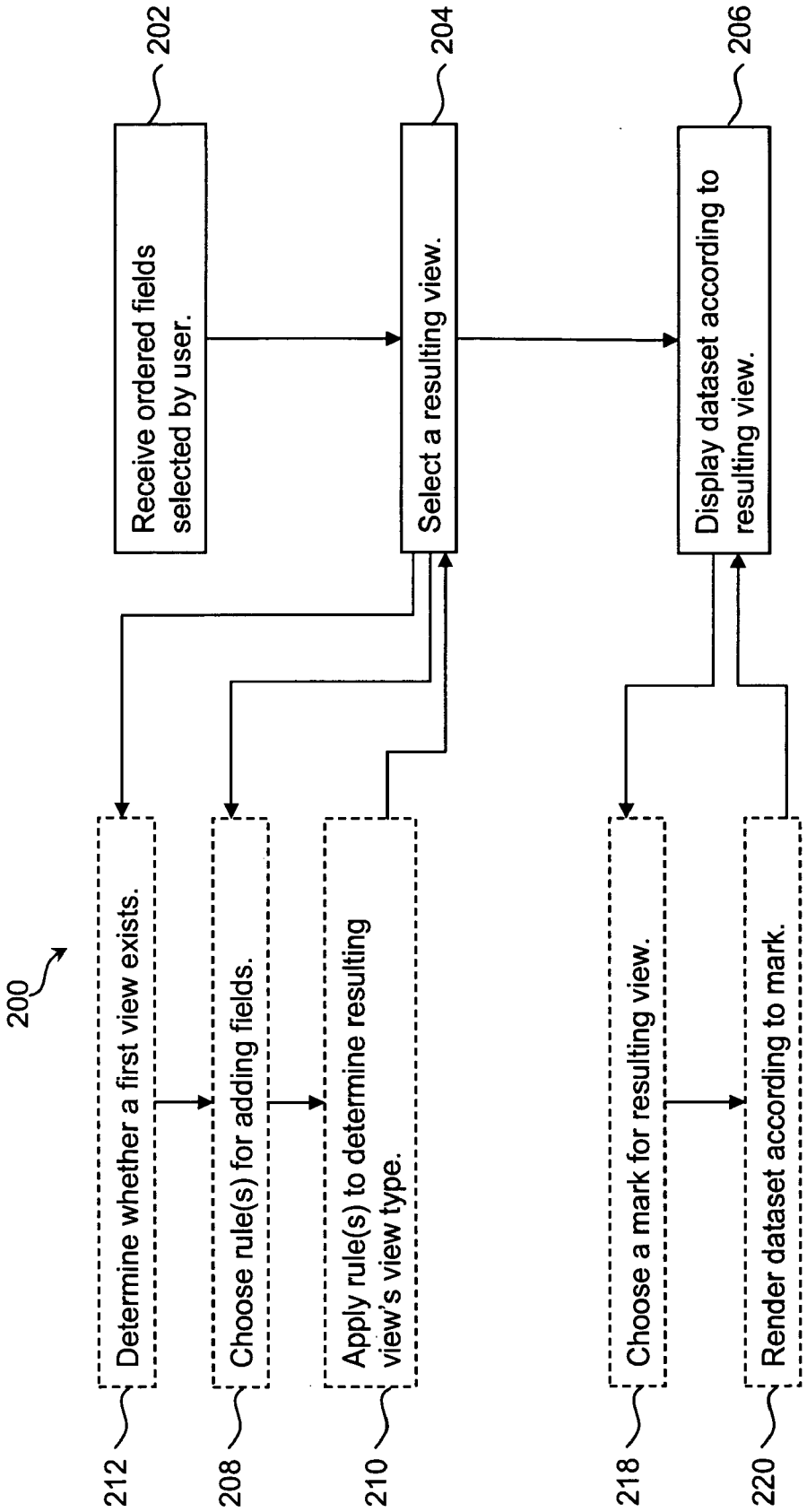

200
Receive ordered fields selected by user.
202
Select a resulting view.
204
Display dataset according to resulting view.
206
Determine whether a first view exists.
212
Choose rule(s) for adding fields.
208
Apply rule(s) to determine resulting view's view type.
210
Choose a mark for resulting view.
218
Render dataset according to mark.
220

Figure 2

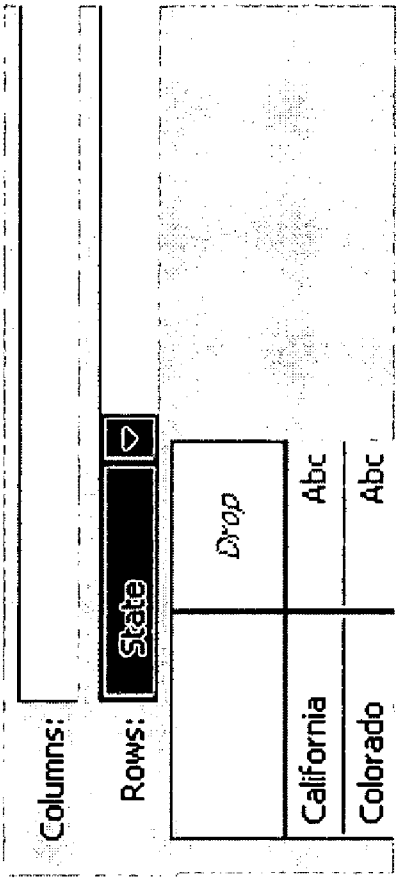


Figure 3b

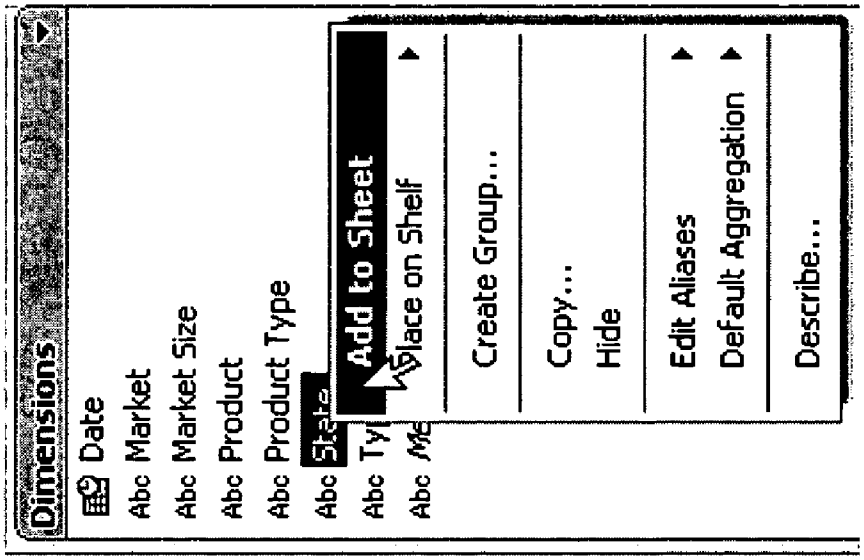


Figure 3a

Drop O

Y/X	ϕ	O	Qi	Qd
ϕ	Y=O	Y=O	Y=O	Y=O
O	X=O	Y+=O	?E+=O Y+=O	?E+=O Y+=O
Qi	X=O	?E+=O X+=O	E+=O	?E+=O X+=O
Qd	X=O	?E+=O X+=O	?E+=O Y+=O	E+=O

Notes

1. E+=O adds to shape then color then Z. The O is always added.
2. ?E+=O conditionally adds to shape then color. O is only added if the encoding is empty. And if the cardinality of the field is less than the cardinality supported by the encoding.
3. In the above rules, the shape shelf is only considered if the mark is shape.
4. Nothing is dropped on an encoding if the mark is a bar. That is, the guarded rule always fails. We try to avoid stacking of bars (more generally, nothing is dropped if we are stacking because we don't know if the field supports stacking (is additive)).
5. The Qi row represents Qi on the Y axis. These cases are rare because Qi is normally placed on the X axis.

Figure 4a

Drop Qd on Pane

Y/X	ϕ	O	Qi	Qd
ϕ	?E*=Qd	?E*=Qd		
O	?E*=Qd	?E*=Qd		
Qi			?E*=Qd	
Qd				?E*=Qd

Notes

1. This set of rules is used when dropping a Qd on a pane. If no match is found, then we use the more general rules for dropping a Qd. The goal is to try to add the Qd to the encodings when Qd is dropped on the pane.
2. The guarded version of the *= operator only adds Qd to text or size. If it can't, that operator is not applied.

Figure 4c

Drop Qd

Y/X	ϕ	O	Qi	Qd
ϕ	E=ϕ?Y=Qd E*=Qd	E*=Qd	Y=Qd	Y=Qd
O	E*=Qd	E*=Qd	Y+=Qd	X+=Qd
Qi	X=Qd	X+=Qd	E*=Qd	X+=Qd
Qd	X=Qd	Y+=Qd	Y+=Qd	XY+=Qd

	ϕ	Ot/O
ϕ		Y+=Qd
O/ Ot	X+=Qd	Y=Ot?X+=Qd Y+=Qd

Notes

1. The guard E=ϕ? checks whether the encoding (text or size) is empty.
2. E*=Qd first tries to add to text then size then color then Z.
3. ?E*=Qd first tries to add to text then size..
4. In the above rules, the text shelf is only considered if the mark is text.
5. An encoding accepts the Qd if it is empty or already contains a Qd. It does not accept the Qd if it contains an O.
6. If the field accepting the encoding contains a Qd or a Qm, then the new Qd is blended with the contents to produce a Qm. If an Om is generated by the blend, it is added to the column shelf if that shelf is empty, otherwise it is added to the row shelf. (Might want to think about this some more?)
7. The Ot rules take precedence over the O rules.
8. The guard Y=Ot? checks whether the Y shelf contains an Ot.
9. The Ot rows represents Ot on the Y axis. This should be rare since best practices dictate that Ot should be on the X axis.
10. XY+=Qd converts to a matrix of scatterplots. All Qds on the rows and columns are the same.

Conversions

1. The ϕϕ rules will drop Qd on the Y axis, unless the text or size encoding contains a Qd. That is, if we are building a text table, continue to build a text table; otherwise make a chart.
2. The OO rules continue to add measures to the text table.
3. The Qiϕ and QiO rules creates QiQd (line) graphs.
4. The ϕQd rules create QdQd scatterplots
5. The OQd rules create stacked charts.
6. The QdQd rule creates a matrix of scatterplots.

Figure 4b

Drop Qi

Y?X	*
*	X!=Qd?X÷=Qd

Notes

1. Always drop Qi on the X axis.
2. But don't do it if the X axis contains a Qd.
3. If there is an existing Qi, then the Qi's are concatenated

Issues

1. Is there something else reasonable to do it the X axis contains a Qd?

Figure 4d

Drop Ot

Y/X	*
*	X÷=Qd

Notes

1. Always drop Ot the X axis.
2. If there is an existing Qi or Qd, then move the Ot leftward until all the Q's are to the right.

Figure 4e

Drop Ot on Pane

Y/X	*
*	E+=Ot

Figure 4f

Tentative Drop Od

Y/X	*
*	E+=Od

Notes

1. The += operator only adds Od to shape or color. If it can't, that operator is not applied.
2. If no match is found, then we don't add the Od. We really want to put Od on an encoding.

Figure 4g

Tentative Drop Qm

Y/X	ϕ|O
ϕ|O	Y=Qm

Figure 4h

Tentative Drop Om

Y/X	ϕ|O
ϕ|O	E=ϕ?E=Qm

Notes:

1. Do nothing if a Q is already being used.
2. Apple Qd rules if one of the O's is an Ot.
3. Adding a Qd makes a bar chart.
4. Adding a Om makes a text table.

Figure 4i

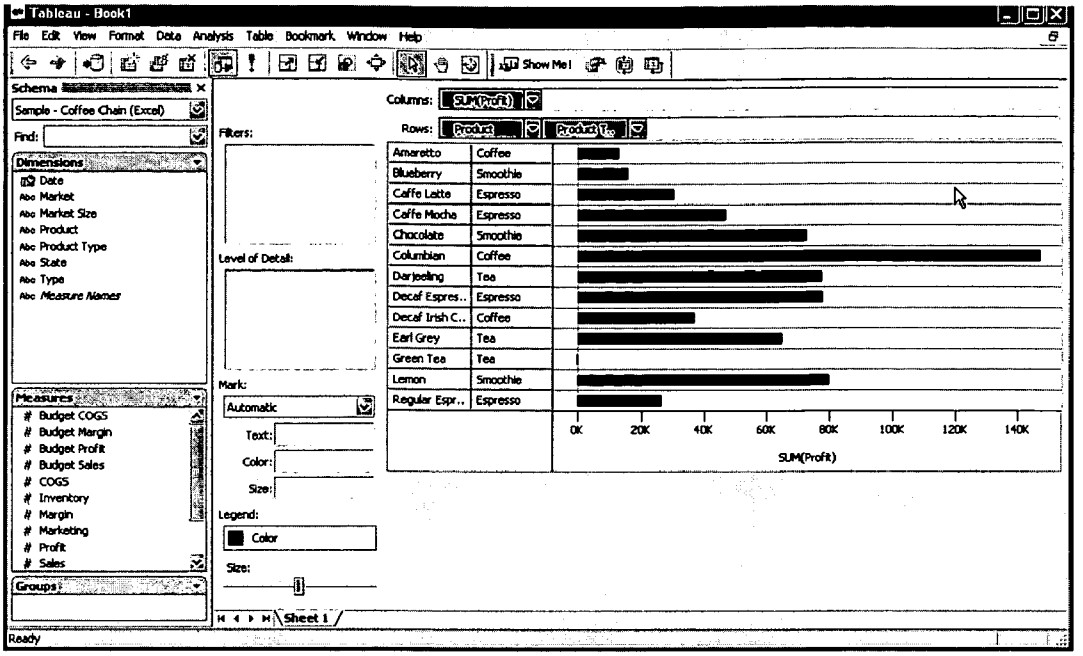


Figure 5a

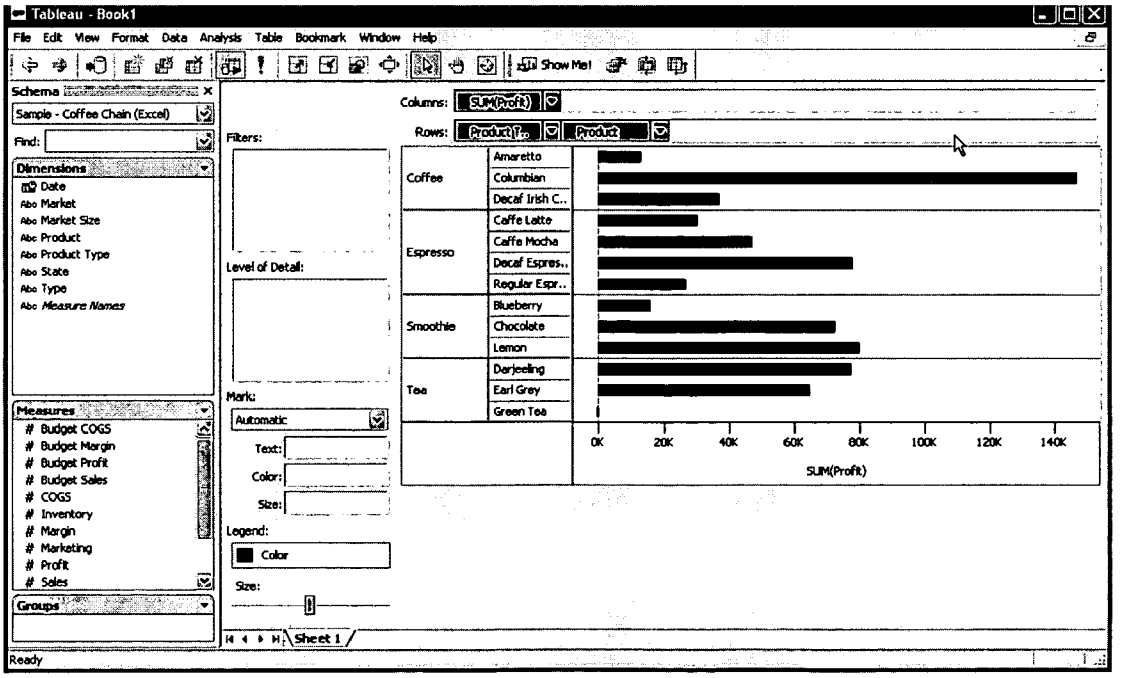


Figure 5b

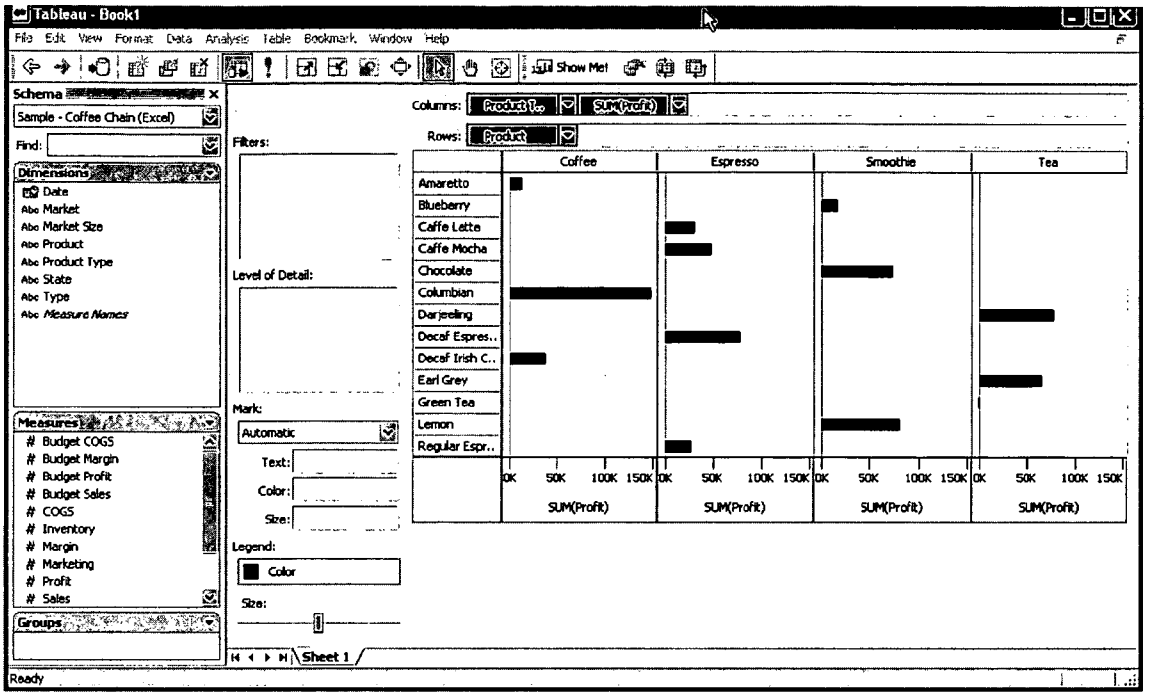


Tableau - Book1
File Edit View Format Data Analysis Table Bookmark Window Help
Show Me!
Schema
Sample - Coffee Chain (Excel)
Find:
Dimensions
Date
Abc Market
Abc Market Size
Abc Product
Abc Product Type
Abc State
Abc Type
Abc Measure Names
Measures
Budget COGS
Budget Margin
Budget Profit
Budget Sales
COGS
Inventory
Margin
Marketing
Profit
Sales
Groups
Filters:
Level of Detail:
Mark:
Automatic
Text:
Color:
Size:
Legend:
Color
Size:
Columns:
SUM(Profit)
Rows:
Product
Coffee
Espresso
Smoothie
Tea
Amaretto
Blueberry
Caffe Latte
Caffe Mocha
Chocolate
Columbian
Darjeeling
Decaf Espres..
Decaf Irish C..
Earl Grey
Green Tea
Lemon
Regular Espr..
0K 50K 100K 150K
SUM(Profit)
Sheet 1
Ready

Figure 5c

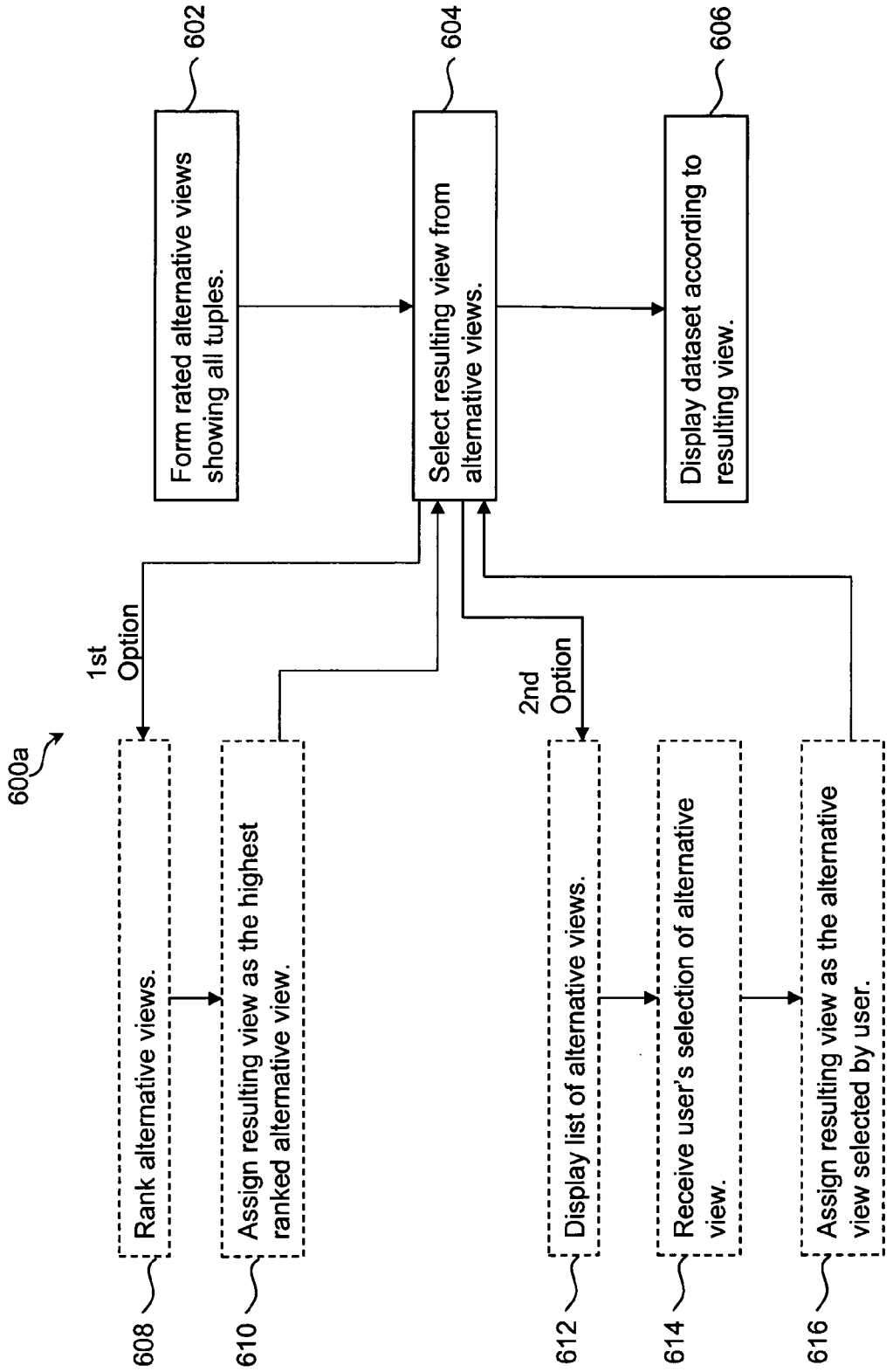

600a
602
Form rated alternative views showing all tuples.
604
Select resulting view from alternative views.
606
Display dataset according to resulting view.
1st Option
2nd Option
608
Rank alternative views.
610
Assign resulting view as the highest ranked alternative view.
612
Display list of alternative views.
614
Receive user's selection of alternative view.
616
Assign resulting view as the alternative view selected by user.

Figure 6a

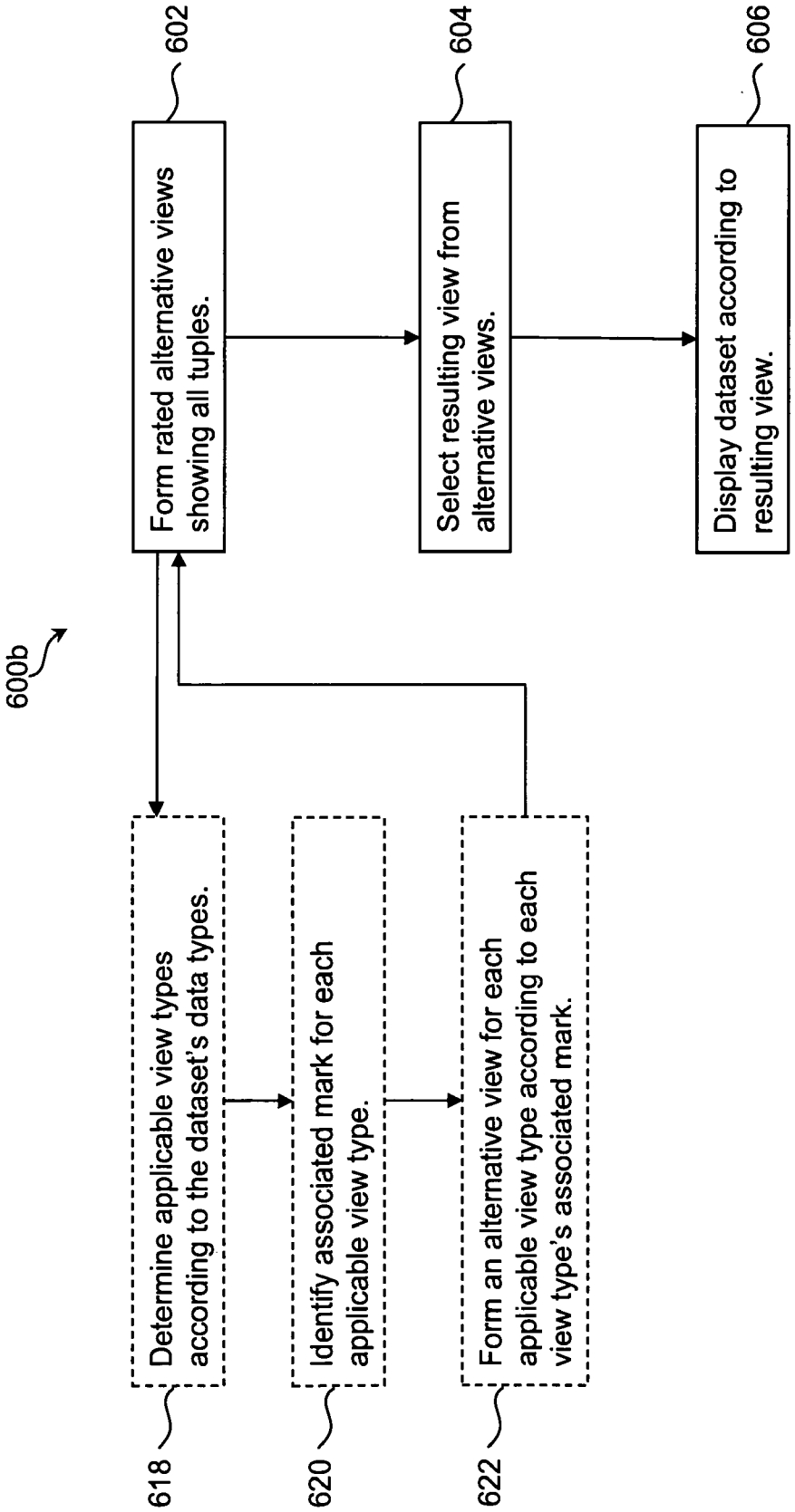

600b
602
Form rated alternative views showing all tuples.
604
Select resulting view from alternative views.
606
Display dataset according to resulting view.
618
Determine applicable view types according to the dataset's data types.
620
Identify associated mark for each applicable view type.
622
Form an alternative view for each applicable view type according to each view type's associated mark.

Figure 6b

	Data	Agg.	O	Qd	Qi	Dates	Rating
Text Table		Yes			No		1
Heat Map		Yes	Yes	1 to 2	No		0
Gantt		Yes	Yes	< 3	Qi or relational date		10 if No Date
Bar (Side-by-side)		Yes	Yes	Yes	No		2
Bar (Stacked)		Yes	>= 2	Yes	No		3 if O >=3
Bar (Measure)		Yes	Yes	>= 2	No		0
Line (Dimension)		Yes		Yes	No	Yes	4
Line (Measure)		Yes		Yes	Qi or relational date		9 if No Date
Circle			Yes	Yes	No		0
Scatter (Single)				2 to 4	No		6
Scatter (Matrix)				>= 3	No		7 if O <= 6
Histogram	Relational	Yes		= 1	No		

Figure 7

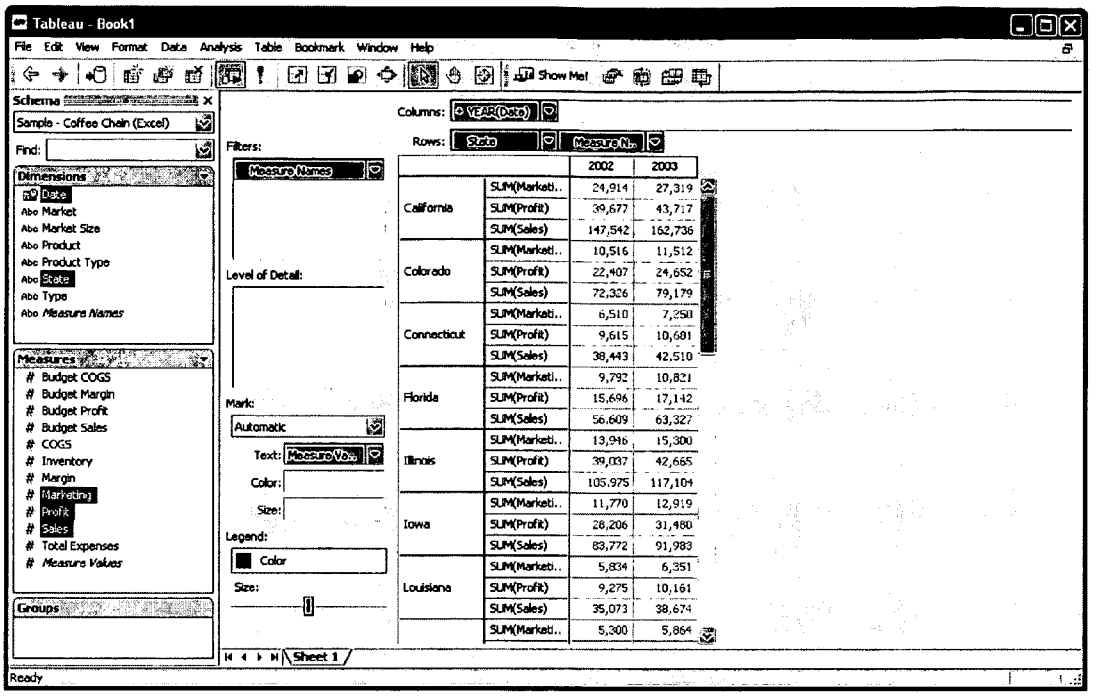


Figure 8a
(Text Table)

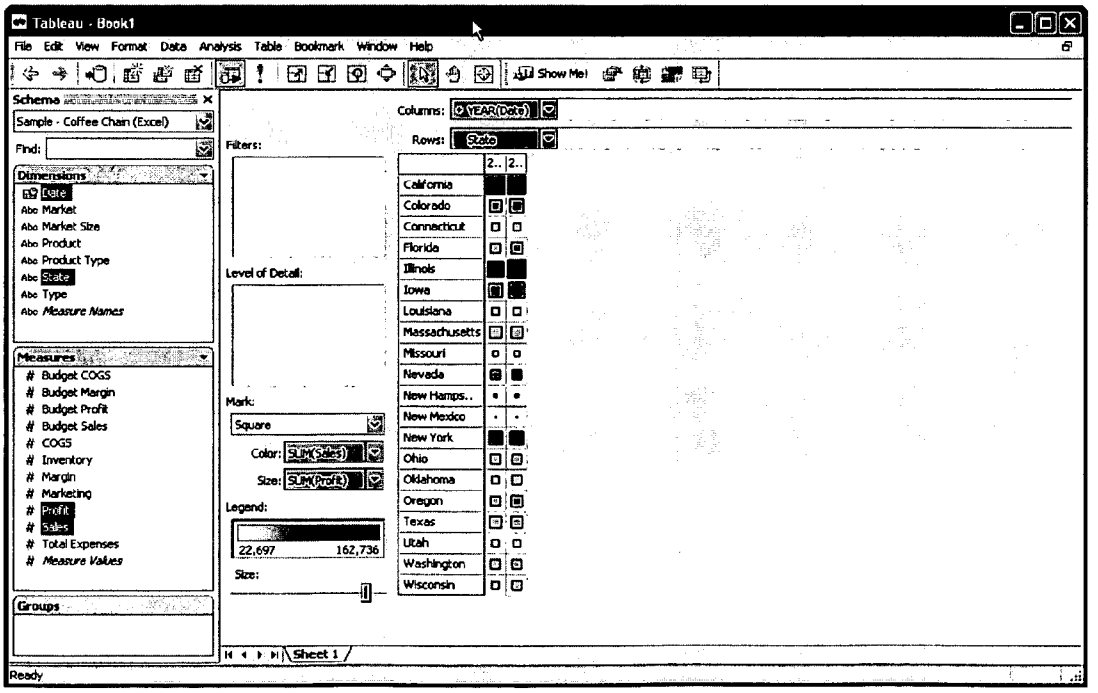


Figure 8b
(Heat Map)

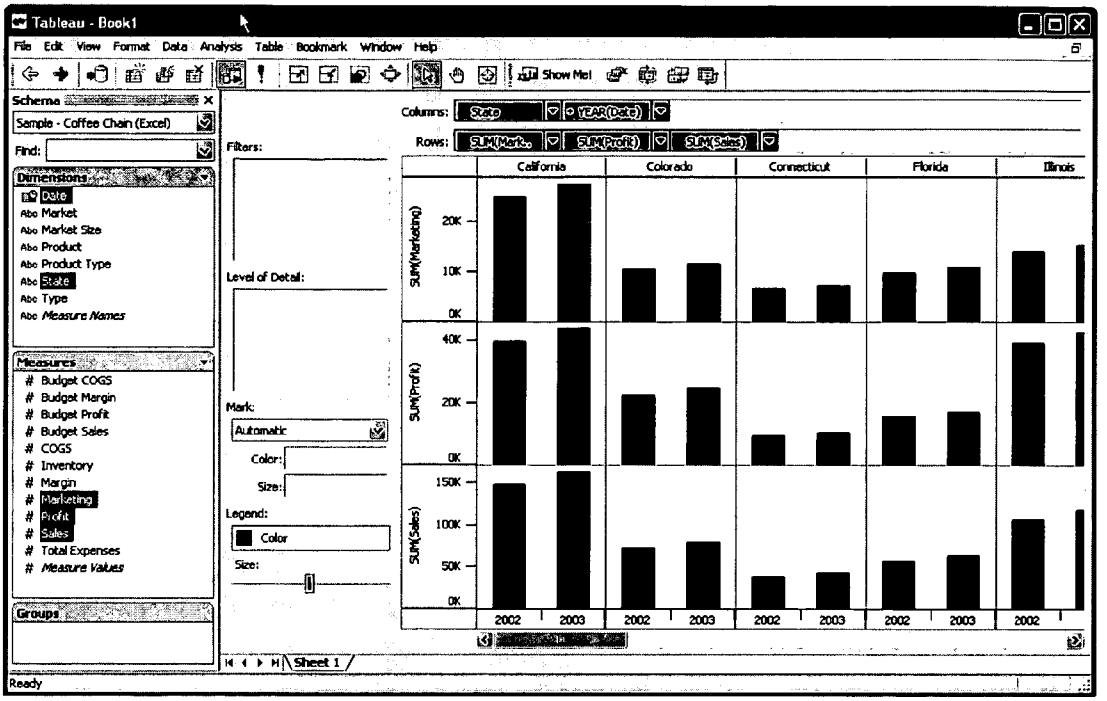


Figure 8c
(Side-By-Side Bar)

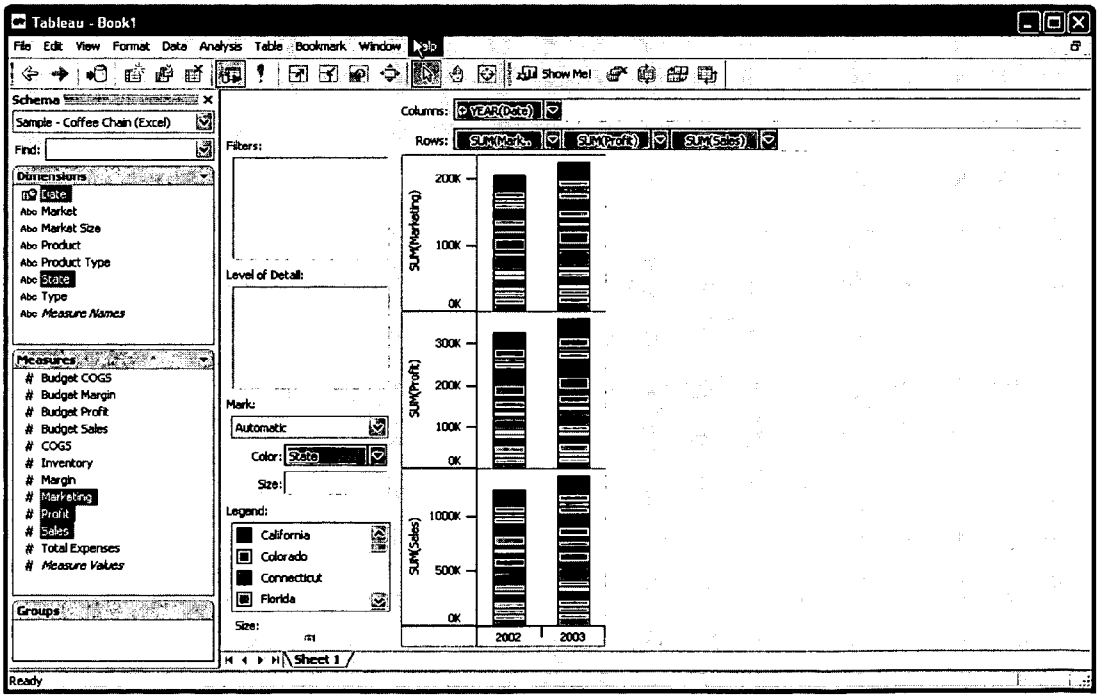


Figure 8d
(Stacked Bar)

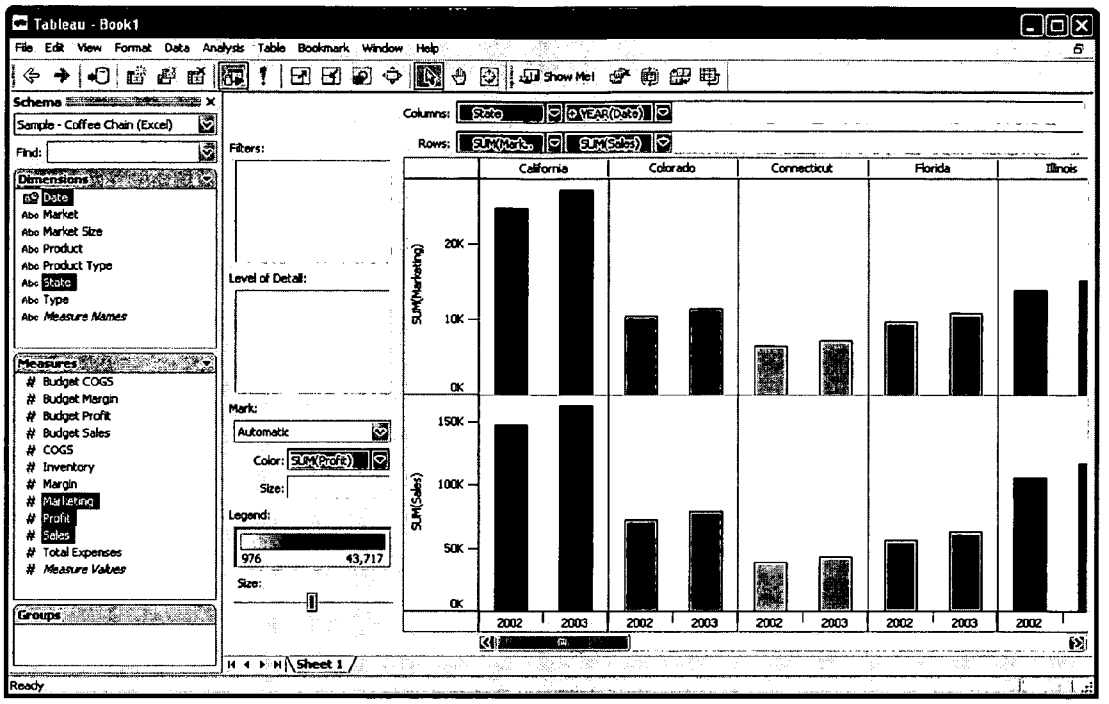


Figure 8e
(Measure Bar)

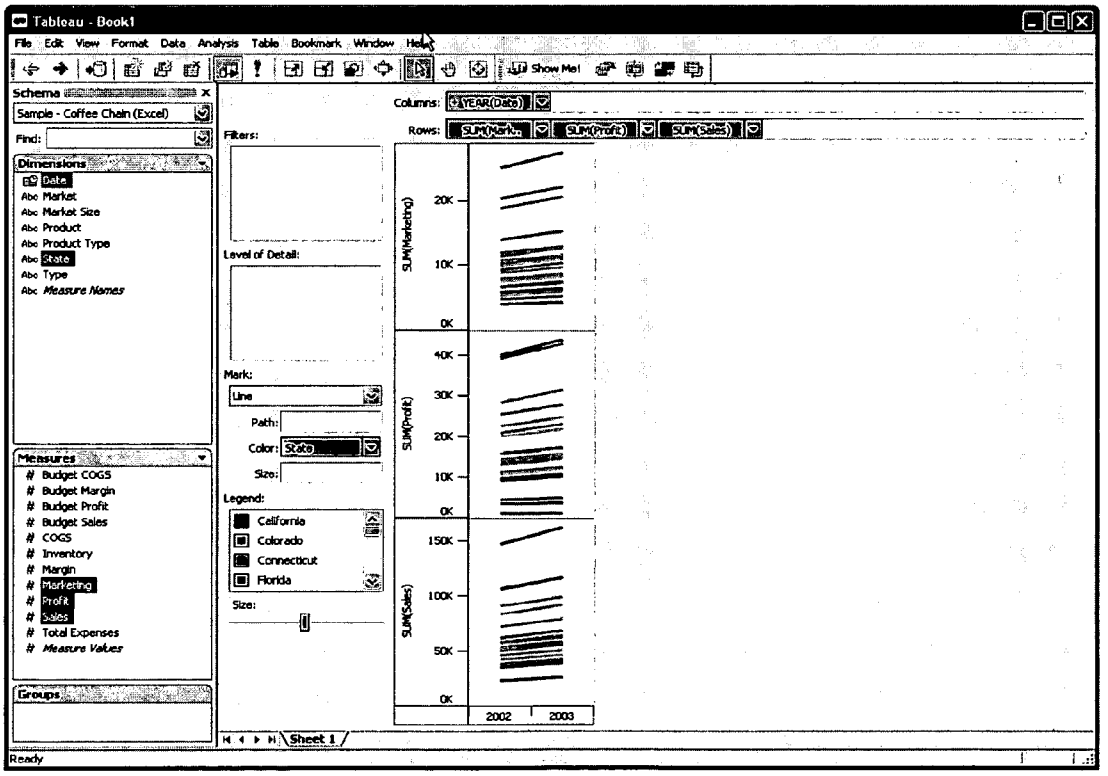


Figure 8f
(Dimension Line)

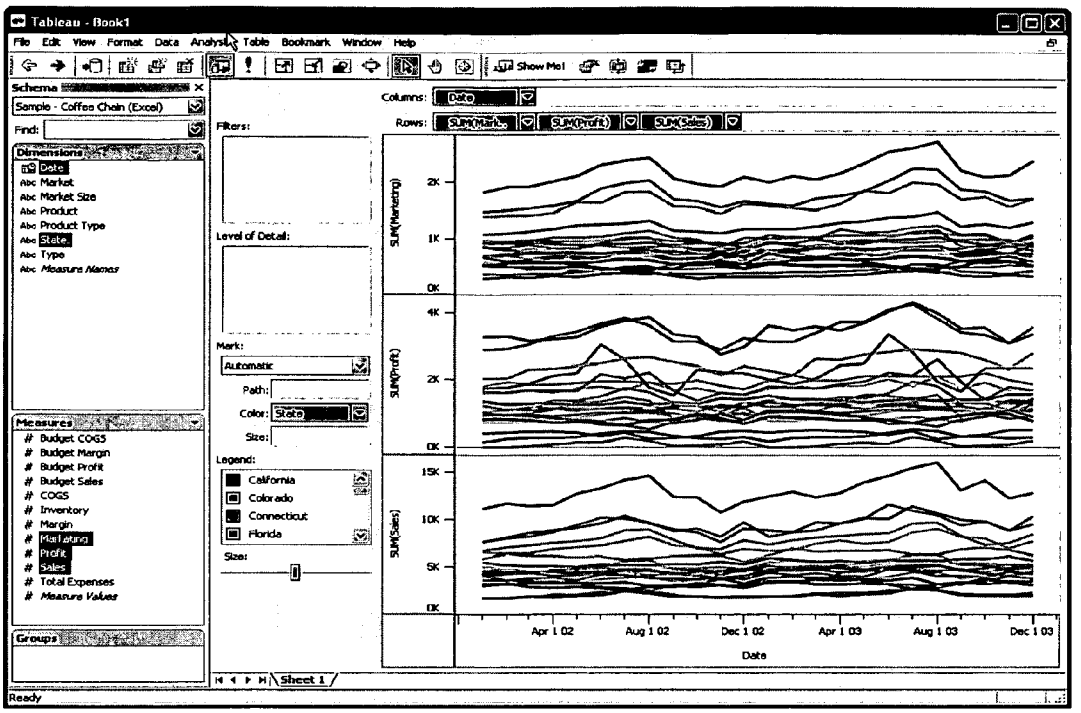


Figure 8g
(Measure Line)

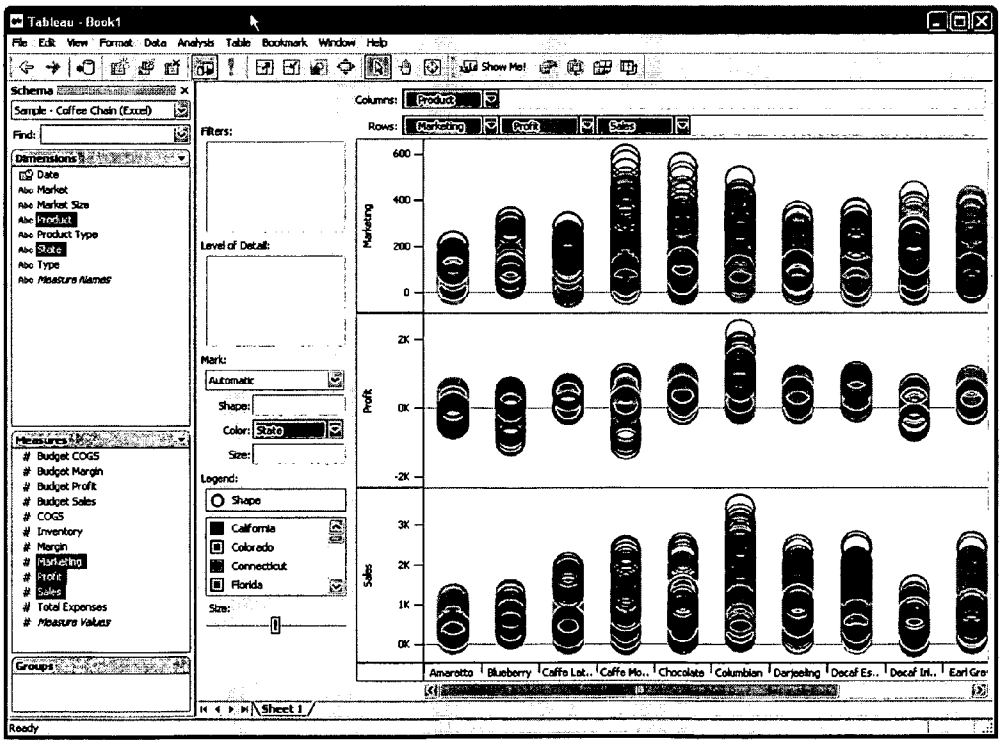


Figure 8h
(Dimension Line)

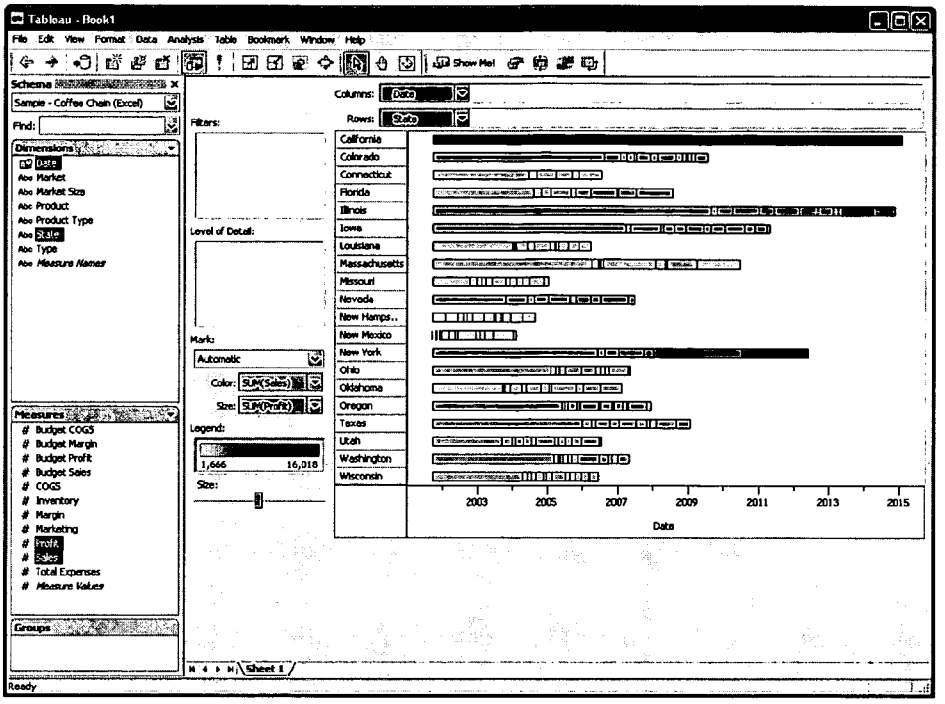


Figure 8i
(Gantt)

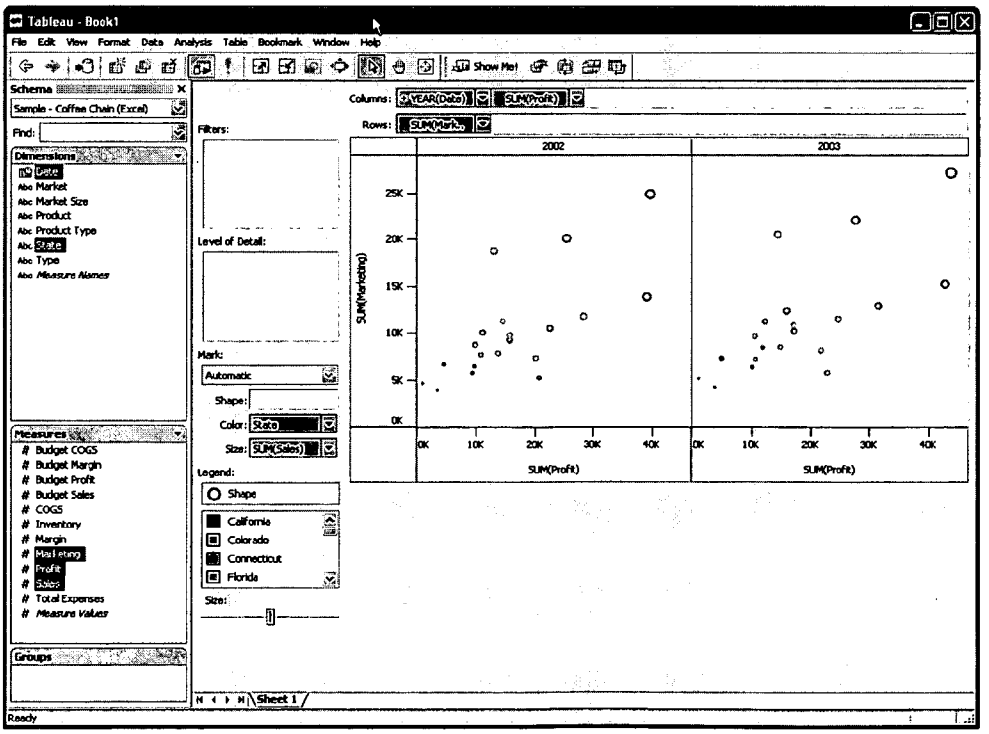


Figure 8j
(Single Scatter)

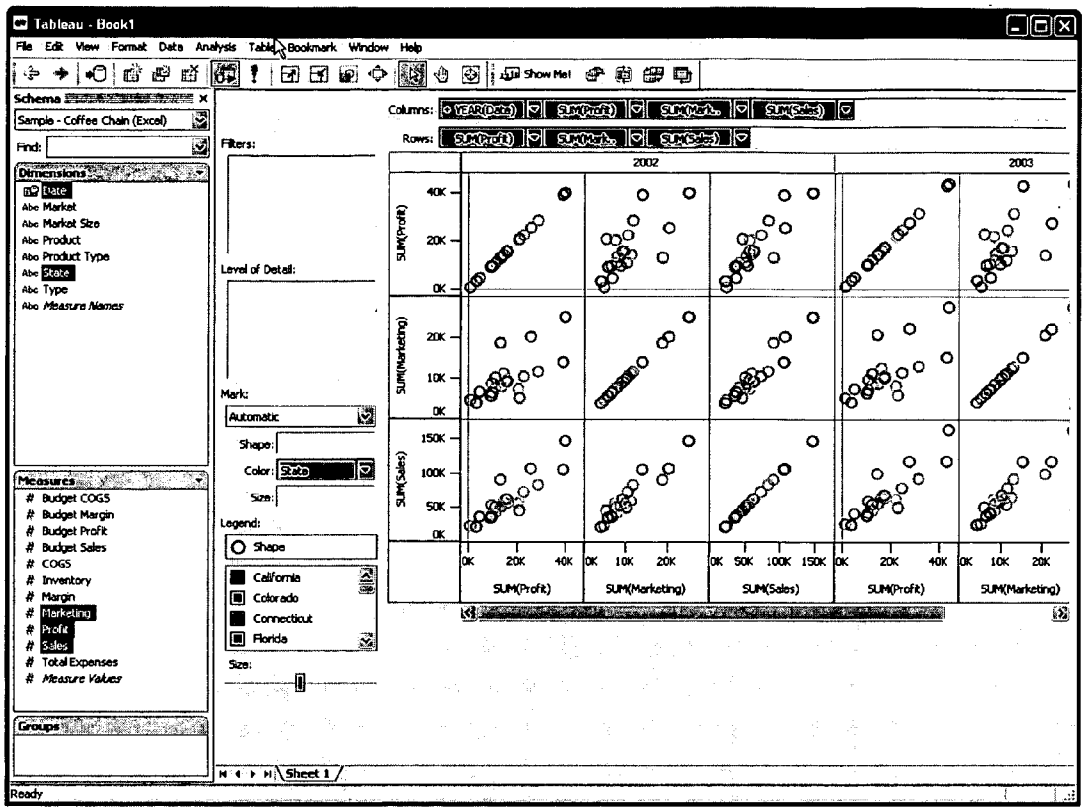


Figure 8k
(Scatter Matrix)

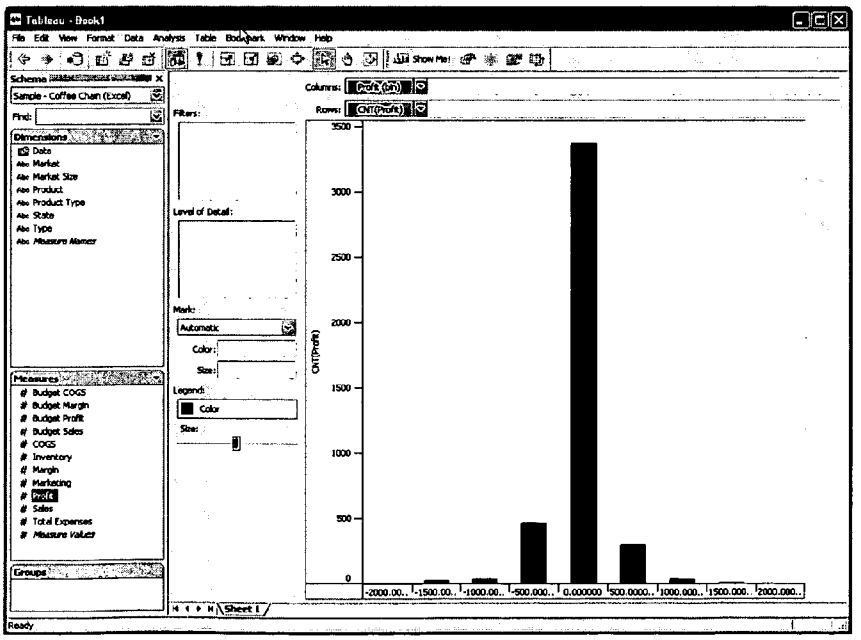


Figure 8l
(Histogram)

COMPUTER SYSTEMS AND METHODS FOR AUTOMATICALLY VIEWING MULTIDIMENSIONAL DATABASES

FIELD OF THE INVENTION

This invention relates generally to computer systems and methods for displaying data such as database information. The invention relates specifically to a computer system and method for displaying data clearly and effectively based upon the types of data found in a dataset.

BACKGROUND OF THE INVENTION

Data is more than the numbers, values, or predicates of which it is comprised. Data resides in multi-dimensional spaces which harbor rich and variegated landscapes that are not only strange and convoluted, but are not readily comprehendible by the human brain. The most complicated data arises from measurements or calculations that depend on many apparently independent variables. Data sets with hundreds of variables arise today in many contexts, including, for example: gene expression data for uncovering the link between the genome and the various proteins for which it codes; demographic and consumer profiling data for capturing underlying sociological and economic trends; sales and marketing data for huge numbers of products in vast and ever-changing marketplaces; and environmental measurements for understanding phenomena such as pollution, meteorological changes and resource impact issues. International research projects such as the Human Genome Project and the Sloan Digital Sky Survey are also forming massive scientific databases. Furthermore, corporations are creating large data warehouses of historical data on key aspects of their operations. Corporations are also using desktop applications to create many small databases for examining specific aspects of their business.

One challenge with any of these databases is the extraction of meaning from the data they contain: to discover structure, find patterns, and derive causal relationships. Often, the sheer size of these data sets complicates this task and means that interactive calculations that require visiting each record are not plausible. It may also be infeasible for an analyst to reason about or view the entire data set at its finest level of detail. Even when the data sets are small, however, their complexity often makes it difficult to glean meaning without aggregating the data or creating simplifying summaries.

Among the principal operations that may be carried out on data, such as regression, clustering, summarization, dependency modeling, and classification, the ability to see patterns rapidly is of paramount importance. Data comes in many forms, and the most appropriate way to display data in one form may not be the best for another. In the past, where it has been recognized that many methods of display are possible, it has been a painstaking exercise to select the most appropriate one. However, identifying the most telling methods of display can be intimately connected to identifying the underlying structure of the data itself.

Business intelligence is one rapidly growing area that benefits considerably from tools for interactive visualization of multi-dimensional databases. A number of approaches to visualizing such information are known in the art. However, although software programs that implement such approaches are useful, they are often unsatisfactory. Such programs have interfaces that require the user to select the most appropriate way to display the information.

Visualization is a powerful tool for exploring large data, both by itself and coupled with data mining algorithms. However, the task of effectively visualizing large databases imposes significant demands on the human-computer interface to the visualization system. The exploratory process is one of hypothesis, experiment, and discovery. The path of exploration is unpredictable, and analysts need to be able to easily change both the data being displayed and its visual representation. Furthermore, the analyst should be able to first reason about the data at a high level of abstraction, and then rapidly drill down to explore data of interest at a greater level of detail. Thus, a good interface both exposes the underlying hierarchical structure of the data and supports rapid refinement of the visualization.

Currently, Tableau's software and Microsoft's Excel are examples of visualization software that create views of datasets. Specifically, Tableau Table Drop allows users to drag data fields onto a Tableau view to create a graphical views. When the view was a text table, the behavior was similar to the drags supported by Excel Pivot Tables. For example, dragging a quantitative data type (Q) onto a text table (X=O Y=O T=Q, where "O" stands for ordinal data), would extend the table to put the two measures next to each other (X=O Y=O,Om T=Qm, where "Om" stands for measure ordinal data and "Qm" stands for measure quantitative data). However, Tableu's Table Drop has functionality not found in Excel's Pivot Tables in that it may change the view type of a view when fields are dragged onto the view. For example, dragging a Q onto a bar chart (X=O Y=Q) created a stacked bar chart (X=O Y=Qm C=Om). Or, if there was already a field with a color encoding (X=O Y=Q C=F) in the view, then the software would transform the Q data into Qm data, and would place the measure names on the Level of Detail encoding (X=O Y=Qm C=F L=Om). With scatter plots, the logic was similar, except the transformation of Q to Qm and placement of measure names on the Level of Detail encoding would be triggered if an existing field already had a shape encoding.

In addition to various software programs, the known art further provides formal graphical presentations. Bertin's Semiology of Graphics, University of Wisconsin Press, Madison Wis. (1983), is an early attempt at formalizing graphic techniques. Bertin developed a vocabulary for describing data and techniques for encoding the data into a graphic. Bertin identified retinal variables (position, color, size, etc.) in which data can be encoded. Cleveland (The Elements of Graphing Data, Wadsworth Advanced Books and Software, (1985), Pacific Grove, Calif. and Visualizing Data, (1993), Hobart Press) used theoretical and experimental results to determine how well people can use these different retinal properties to compare quantitative variations.

Mackinlay's APT system (ACM Trans. Graphics, 5, 110-141, (1986)) was one of the first applications of formal graphical specifications to computer generated displays. APT uses a graphical language and a hierarchy of composition rules that are searched through in order to generate two-dimensional displays of relational data. The Sage system (Roth, et al., (1994), Proc. SIGCHI '94, 112-117) extends the concepts of APT, providing a richer set of data characterizations and forming a wider range of displays. The existing art also provides for the assignment of a mark based upon the innermost data column and row of a dataset (Hanrahan, et al., U.S. Pat. application Ser. No. 11/005,652, "Computer System and Methods for Visualizing Data with Generation of Marks"). Heuristically guided searches have also been used to generate visualizations of data (Agrawala, et al., U.S. Pat. No. 6,424,933, "System and Method for Non-Uniform Scaled Mapping").

A drawback with the formal graphical specifications of the art is that they do not provide any guidance to a user as to useful and clear visual formats in which a set of data could be rendered. The rendering of the data is such that there is no analysis to examine the resulting visualization for clarity or usefulness. Further, in the use of heuristic searches (trial-and-error method), the searches fail, leaving the user with the problem of finding clear or useful views. Heuristic algorithms can have complex behavior that creates a poor user experience. When a user does not understand why a heuristic algorithm generates certain views, the algorithm becomes unpredictable to the user and the user will not be inclined to use the algorithm.

Based on the background state of the art, as described herein, what is needed are improved methods and graphical interfaces wherein the initial visualization of data has been determined to be a clear and useful visualization, and this visualization is then automatically presented to the user.

SUMMARY OF THE INVENTION

The present invention provides improved methods for visualizing data.

A first aspect of the invention provides a computer implemented method for automatically and visually displaying a graphical representation of a dataset, comprising: receiving a user selected and ordered plurality of fields; selecting a resulting view for displaying the dataset based on the order of the user selected fields; and displaying the dataset or a transformation of the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view.

A second aspect of the invention provides a computer implemented method for automatically and visually displaying a graphical representation of a dataset with a plurality of tuples, comprising: forming a plurality of rated alternative views, each alternative view showing all tuples, or a transformation of all tuples, in the dataset; selecting a resulting view from the plurality of alternative views, based upon a user selected option; and displaying the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view. In yet a further embodiment, when the user selected option is a first option, the selecting step further comprises: ranking the plurality of alternative views according to a rating system; and assigning the resulting view as the highest ranked alternative view. In still another embodiment, when the user selected option is a second option, the selecting step further comprises: displaying a list of the alternative views; receiving the user's selection of an alternative view; and assigning the resulting view as the alternative view selected by the user.

A third aspect of the invention provides a computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism for automatically and visually displaying a graphical representation of a dataset, the computer program mechanism comprising: a field receiver for receiving a user selected and ordered plurality of fields; a resulting view selector for selecting a resulting view for displaying the dataset based on the order of the user selected fields; and a dataset displayer for displaying the dataset or a transformation of the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view.

A fourth aspect of the invention provides a computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism for automatically and visually displaying a graphical representation of a dataset with a plurality of tuples, the computer program mechanism comprising: an alternative view former for forming a plurality of rated alternative views, each alternative view showing all tuples, or a transformation of all tuples, in the dataset; a resulting view selector for selecting a resulting view from the plurality of alternative views, based upon a user selected option; and a dataset displayer for displaying the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view. In yet a further embodiment, when the user selected option is a first option, the resulting view selector further comprises: an alternative view ranker for ranking the plurality of alternative views according to a rating system; and a view assignor for assigning the resulting view as the highest ranked alternative view. In still another embodiment, when the user selected option is a second option, the resulting view selector further comprises: a list displayer for displaying a list of the alternative views; a selection receiver for receiving the user's selection of an alternative view; and a view assignor for assigning the resulting view as the alternative view selected by the user.

A fifth aspect of the invention provides a computer system for automatically and visually displaying a graphical representation of a dataset, the computer system comprising: a central processing unit; a memory, coupled to the central processing unit, the memory storing: the dataset; a programming module comprising, comprising: instructions for receiving a user selected and ordered plurality of fields; instructions for selecting a resulting view for displaying the dataset based on the order of the user selected fields; and instructions for displaying the dataset or a transformation of the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view.

A sixth aspect of the invention provides a computer system for automatically and visually displaying a graphical representation of a dataset with a plurality of tuples, the computer system comprising: a central processing unit; a memory, coupled to the central processing unit, the memory storing: the dataset; a programming module comprising: instructions for forming a plurality of rated alternative views, each alternative view showing all tuples, or a transformation of all tuples, in the dataset; instructions for selecting a resulting

view from the plurality of alternative views, based upon a user selected option; and instructions for displaying the dataset according to the resulting view. In one embodiment, the dataset is retrieved from a remote database. In another embodiment, rules are used to select the resulting view. In yet another embodiment, the rules are predetermined. In other embodiments, the rules are determined by the user's preferences or usage. In a further embodiment, heuristics are used to select the resulting view. In yet a further embodiment, when the user selected option is a first option, the instructions for selecting further comprises: instructions for ranking the plurality of alternative views according to a rating system; and instructions for assigning the resulting view as the highest ranked alternative view. In still another embodiment, when the user selected option is a second option, the instructions for selecting further comprises: instructions for displaying a list of the alternative views; instructions for receiving the user's selection of an alternative view; and instructions for assigning the resulting view as the alternative view selected by the user.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is an illustration of a computer system that facilitates the visualization of a dataset in a clear and useful form.

FIG. **2** is a flowchart of the steps through which a system proceeds in one embodiment of the present invention.

FIGS. **3***a* and **3***b* are illustrations of one way of presenting an embodiment of the present invention to a user.

FIGS. **4***a*-**4***i* are rules whereby a field may be added to an existing view or whereby a field may be used as the only field in a view.

FIGS. **5***a*-**5***c* are examples of different views of a single dataset.

FIGS. **6***a* and **6***b* are flowcharts of the steps through which a system proceeds in another embodiment of the present invention.

FIG. **7** is a table showing the criteria for forming views of a dataset and a rating system for one embodiment of the present invention.

FIGS. **8***a*-**8***l* are resulting views, based upon an embodiment of the present invention.

Like reference numerals refer to corresponding parts throughout the several views of the drawings.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods, computer program products, and computer systems for automatically providing a user with a clear and useful view of a dataset. In a typical embodiment, the present invention builds and displays a view of a dataset as a user adds fields to the dataset or as a dataset is accessed, such that the view is clear and useful, and is automatically presented to the user. An advantage of the present invention is that data is presented in a clear and useful form automatically.

The present invention operates on a set of data, called a dataset, that are made up of tuples. As one skilled in the art will realize, the dataset can be a relational database, a multidimensional database, a semantic abstraction of a relational database, or an aggregated or unaggregated subset of a relational database, multidimensional database, or semantic abstraction. Fields are categorizations of data in a dataset. A tuple is an item of data (such as a record) from a dataset, specified by attributes from fields in the dataset. A search query across the dataset will return one or more tuples. Fields contain data that are of particular types, and each field is of a particular type. These types include:

Data Type	Symbol
Ordinal	O
Ordinal time (date)	Ot
Dependent ordinal (categorical measure)	Od
Measure names	Om
Quantitative	Q
Independent Quantitative (dimension)	Qi
Dependent Quantitative (measure)	Qd
Measure values	Qm
Quantitative time	Qt
Quantitative position	Qx

Measure names may include an ordinal field whose domain is the name of one or more Qd fields. Measure values may include a dependent quantitative field whose domain and values are the blending of the Qd fields whose names appear in the domain of measure names.

A view is a visual representation of a dataset or a transformation of that dataset. Text table, bar chart, and scatter plots are all examples of types of views. Views contain marks that represent one or more tuples in a dataset. In other words, marks are visual representations of tuples in a view. A mark is typically associated with a type of graphical display. Some examples of views and their associated marks are as follows:

View Type	Associated Mark
Table	Text
Scatter Plot	Shape
Bar Chart	Bar
Gantt Plot	Bar
Line Graph	Line Segment
Circle Graph	Circle

FIG. **1** is an illustration of a computer system that facilitates the visualization of a dataset in a clear and useful form. System **100** includes memory **102**, CPU **180**, user interface **184**, storage unit **194**, disk controller **192**, and bus **182** that connects all of system **100**'s elements together. System **100** may also have network connection **196** for communication with other systems on a network. System **100** also includes memory **102**, which stores operating system **104**, file system **106**, as well as various other modules related to the present invention. Additionally, memory **102** may also store dataset **140**, which contains tuples. User interface **184** further includes display **186**, mouse **188**, and keyboard **190**. System **100** may also be connected to database **150** where a dataset may be retrieved and stored in memory **102**. Memory **102** also stores computer program mechanisms that are necessary to some embodiments of the present invention.

In FIG. **2**, flowchart **200** describes the steps through which a system proceeds in one embodiment of the invention. At step **202**, ordered fields selected by a user are received. A resulting view is selected at step **204**, and the dataset is displayed at step **206** according to the resulting view.

The computer system modules used to perform this embodiment of the invention are shown in FIG. **1**. Field receiver **108** performs step **202** by receiving ordered fields selected by the user. Resulting view selector **110** performs step **204** and selects a resulting view. Dataset displayer **112** performs step **206** and displays the dataset according to the resulting view.

According to one embodiment of the invention, resulting view selector **110** selects the resulting view by choosing rule(s) for adding the user selected ordered fields (step **208**). This is accomplished by rule chooser **114**. Rule applier **116** then applies the rule(s) to determine the resulting view's view type (step **210**). In another embodiment of the invention, before rule chooser **114** chooses rule(s), view determiner **118** determines whether a first view exists (step **212**). In yet another embodiment of the invention, the dataset is displayed in step **206** when mark chooser **126** chooses a mark for the resulting view (step **218**), and dataset renderer **128** renders the dataset according to the mark (step **220**).

FIGS. **4***a*-**4***i* show sets of rules that are associated with adding (or "dropping") fields with particular data types. The field may be the only field in a view, or the field may be in addition to fields already in an existing view. When dropping a field, the field is added either as a column or a row, or it may be encoded. Encodings include color, size, and shape to represent a value. For example, red may represent all values between 1 and 10. The following convention is used for operators in the rules shown in FIGS. **4***a*-**4***i* ("E" designates encoding):

Operator	Limitations
= Assign field to a clause += Add field to the end of the clause (some rearrangements may occur)	Left hand side is a column or row Right hand side must be O or Qd
*= Blend field with column or row (blend Qd with first E accepting/containing a Qd). The blend will result in Qm being on column or row, and an Om being added to the view.	Right hand side must be Qd
? Guard the action. Only add if the column or row accepts the field and the cardinality of the field is less than the cardinality associated with the column or row.	Unary

The sets of rules are organized first by the type of the field that is dropped (e.g. O or Qd), and then by the type of the view that the field is being dropped onto. The rules are further broken down by the type of the view. The type of a view depends on their innermost row and column. For example, OO is a view with ordinal fields in the row and column; OQ is a view with an ordinal field in the row and a quantitative field in the column; and ϕ is an empty view with no fields. For each type of field being dropped, a rule table is shown containing the rules for each type of view into which the field is being dropped. The columns of the rule tables represent the contents of the innermost field on the column (X), and the rows of the rules table the innermost field on the row (Y).

In step **208**, rule(s) for adding the user selected field's data type are chosen. For example, if a user selected field is an ordinal, then the set of rules in FIG. **4***a* would be used. Or, if a user selected field is an independent quantitative, then the set of rules in FIG. **4***d* would be used. If in step **212** view determiner **118** determines that no first view exists, then rule chooser **114** would choose the rule in row **1** column **1** of FIG. **4***a* as the rule in step **208**. If in step **212** view determiner **118** determines that a first view exists, and the first view contains an independent quantitative field in the innermost column and an independent quantitative field in the innermost row of the first view, then rule chooser **114** would select the rule in row **3** column **3** of FIG. **4***a* as the rule in step **208**. Finally, in step **210**, rule applier **116** applies the rule selected by rule chooser **114**. If no first view exists, then the resulting view will contain a single column (Y=O). If the first view was of the QiQi type, then the resulting view will contain an encoded field (E+=O).

The order in which fields are added affect the view type of the resulting view. For example, if a measure data type field is added to an empty view, and is subsequently followed by a dimension data type field, the resulting view will be a bar chart. However, if a measure data type field is added to an empty view subsequent to a dimension data type field, then the resulting view will be a text table. The resulting view's view type is thusly selected based upon a set of rules. The view type is then assigned to the resulting view and the view is then populated with data from the dataset. In one embodiment, the set of rules are predetermined. In another embodiment, the set of rules are based upon a user's preferences or actual usage. For example, a user may be given the opportunity to designate the best view type for various sequences of the addition of fields to views. Or, after the visual plot is populated and rendered for the user, the user is allowed to choose a different rendering. The user's choice as to the ultimate resulting view, if recorded, may indicate the user's preference for what view type the user considers the clear and/or useful. In yet another embodiment, heuristics may be used instead of a set of rules for selecting a resulting view.

In one embodiment, the cardinalities of the fields in the resulting view are computed and are considered in determining how the user selected fields are added. In set theory, cardinality is the size of a set. In the present invention, cardinality refers to the number of distinct instances that are associated with a field's type. For example, if a field type is "States of America", then the cardinality of such a field would be 50.

In another embodiment, the functional dependency of the fields in the resulting view are computed and are considered in determining how the user selected fields are added. Functional dependency refers to the determination of one field by another field. For example, if one field is of the type "States of America," and a second field is "Inches of Rainfall of the States of America," then the second field depends upon the first. Another example is shown in FIGS. **5***a*-**5***c*. Referring to FIGS. **5***a*-**5***c*, there is a functional dependency from the Product field to the Product Type field because each product has a unique product type. This can be seen in FIG. **5***a* because each product (in the column where the product field resides) has a single product type (in the column where the product type field resides) to its right. When the columns are reversed in FIG. **5***b*, it is apparent that there is no functional dependency from the Product Type field to the Product Field because each product type has multiple products. Finally, when the fields that form a functional dependency are placed in both rows and columns (FIG. **5***c*), the resulting view contains much empty space, which makes the resulting view less effective.

In yet another embodiment, in the application of the selected rule to populate the resulting view with data from the dataset, a mark is chosen for the resulting view's view type and the data from the dataset is rendered according to the mark. This is shown in FIG. **2**, where, in step **218**, mark chooser **126** chooses a mark for the resulting view, and data renderer **128** renders the dataset according to the mark in step **220**.

FIGS. **3***a* and **3***b* show ways of operating an embodiment of the present invention. A user may drag a new field from a list of available fields and drop the field onto a view. Another way of operating an embodiment of the present invention is for the user to double click on a new field from a list of available fields. This automatically adds the new field to an existing view or automatically forms a new view if there is not an existing view. Other ways of adding or dropping fields include double clicking on a field, selecting fields, typing

field names, and creating a specification for a set of fields using statistical analysis, historical analysis, or heuristic algorithms.

Now, referring to FIG. **6***a* a flowchart is provided for the steps through which a system proceeds in another embodiment of the present invention. First, alternative view former **130** forms alternative views of the tuples of the dataset (step **602**). Resulting view selector **110** then selects a resulting view from the alternative views (step **604**). Finally, dataset displayer **112** displays the dataset according to the resulting view (step **606**).

In another embodiment, alternative views are formed based upon a set of criteria. FIG. **7** is a table showing the criteria for forming alternative views of a dataset and a rating system for one embodiment of the present invention. For example, if all the data in a dataset is aggregated and does not contain any independent quantitative data, then one of the possible views is a text table as determined by the first rule. Its rating is 1 meaning that it will only be the highest ranking view if other views such as Line (Measure), which was a higher rating of 9, is not applicable to the selected fields. As one skilled in the art will realize, these ratings could also be based on other criteria such as user preference, usage patterns, and statistical analysis of the data.

In one embodiment, if the user selected a first option, then the alternative views are ranked according to a rating system by alternative view ranker **134** in step **608**. View assignor **120** then assigns the resulting view as the highest ranked alternative view at step **610**. Dataset displayer **112** then displays the dataset according to the resulting view in step **606**. For example, if all the data in a dataset is aggregated and does not contain any independently quantitative data, then alternative views of all the view types listed in FIG. **7** are generated at step **602**. Then, at step **604**, the text table alternative view is selected to be the resulting view, and the dataset is displayed as a text table in step **606**. In another embodiment, in accomplishing step **602** (forming rated alternative views showing all tuples), view determiner **118** determines applicable view types according to the dataset's data types. Mark chooser **126** then identifies an associated mark for each applicable view type at step **620**, which is then used to form alternative views for each applicable view type at step **622**.

In another embodiment, if the user selected a second option, then a list of alternative views would be displayed by list displayer **136** at step **612** for the user's selection. After the user's selection is received at step **614** by selection receiver **138**, the resulting view is assigned as the alternative view that the user selected by view assignor **120** at step **616**, and dataset displayer **112** then displays the dataset according to the resulting view in step **606**.

In yet another embodiment of the invention, cardinality computer **122** computes the cardinality of the fields in the plurality of tuples when forming the alternative views. In a further embodiment, functional dependency computer **124** computes the functional dependency of the fields in the plurality of tuples when forming the alternative views.

FIG. **8***a* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a text table, or the user selected the alternative text table view. The dataset must include only aggregated data and no independently quantitative data.

FIG. **8***b* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a heat map, or the user selected the alternative heat map view. The dataset must include only aggregated data, at least one field of ordinal data, one to two fields of dependant quantitative data, and no independent quantitative data.

FIG. **8***c* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a side-by-side bar chart, or the user selected the alternative side-by-side bar chart view. The dataset must include only aggregated data, at least one field of ordinal data, at least one field of dependant quantitative data, and no independent quantitative data.

FIG. **8***d* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a stacked bar chart, or the user selected the alternative stacked bar chart view. The dataset must include only aggregated data, at least two fields of ordinal data, at least one dependent quantitative data, and no independent quantitative data.

FIG. **8***e* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a measure bar chart, or the user selected the alternative measure bar chart view. The dataset must include only aggregated data, at least one field of ordinal data, at least two fields of dependent quantitative data, and no independent quantitative data.

FIG. **8***f* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a dimension line graph, or the user selected the alternative dimension line graph view. The dataset must include only aggregated data, at least one field of dependent quantitative data, at least one field of dates, and no independent quantitative data.

FIG. **8***g* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a measure line graph, or the user selected the alternative measure line graph view. The dataset must include only aggregated data, at least one field of dependent quantitative data, and at least one field of independent quantitative data or dates.

FIG. **8***h* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a circle graph, or the user selected the alternative circle graph view. The dataset must not include any aggregated data or independent quantitative data, and must include at least one field each of ordinal and dependent quantitative data.

FIG. **8***i* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a Gantt chart, or the user selected the alternative Gantt chart view. The dataset must include only aggregated data, at least one field of ordinal data, less than three fields of dependent quantitative data, and at least one field of independently quantitative data or of relational dates.

FIG. **8***j* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a single scatter plot, or the user selected the alternative single scatter plot view. The dataset must include two to four fields of dependent quantitative data, and at least one field of independent quantitative data.

FIG. **8***k* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a matrix scatter plot, or the user selected the alternative matrix scatter plot view. The dataset must include three to six fields of dependent quantitative data, and at least one field of independent quantitative data.

FIG. **8***l* shows a rendering of the data in a dataset in an resulting view where either the highest ranked view type was a histogram, or the user selected the alternative histogram view. The dataset must include only aggregated and relational data, must have exactly one field of dependent quantitative data, and must have no independent quantitative data.

The present invention not only accepts datasets and databases as inputs, it also accepts views as inputs. A view can be

used to represent a set of fields. Resulting views can also depend on the existing view. For example, rules or operators can take into account the current view to generate a new view that is related to the current view. Also, as one skilled in the art will realize, many other rules are possible, include ones to generate statistical, maps, pie charts, and three dimensional views of data.

The present invention can be implemented as a computer program product that comprises a computer program mechanism embedded in a computer readable storage medium. For instance, the computer program product could contain the program modules shown in FIG. **1**. These program modules may be stored on a CD-ROM, magnetic disk storage product, or any other computer readable data or program storage product. The software modules in the computer program product can also be distributed electronically, via the Internet or otherwise, by transmission of a computer data signal (in which the software modules are embedded) on a carrier wave.

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

All references cited herein are incorporated herein by reference in their entirety and for all purposes to the same extent as if each individual publication or patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

What is claimed:

1. A computer implemented method for displaying a graphical representation of a dataset, comprising:

at a computer having one or more processors and memory storing programs executed by the one or more processors:

displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

receiving a user selection of at least a subset of the multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, each of the user-selected terms having an associated data type; and

in response to the user selection and the user ordering,

forming a plurality of alternative views using the dataset based on the user ordering of the user-selected terms;

determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system;

selecting from among the plurality of alternative views a resulting view for displaying the dataset based on comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

displaying the dataset or a transformation of the dataset in the data visualization region according to the resulting view.

2. The computer implemented method of claim **1**, wherein the selecting step further comprises:

with no further user instruction:

choosing one or more rules that are associated with adding the user selected fields; and

applying the one or more rules to determine the resulting view's view type.

3. The computer implemented method of claim **2**, further comprising, before the choosing step:

determining whether a first view exists;

wherein the one or more rules selected is also based upon the first view, if the first view is determined to exist.

4. The computer implemented method of claim **1**, wherein the dataset is a database.

5. The computer implemented method of claim **4**, wherein the database is remote.

6. The computer implemented method of claim **4**, wherein the database is a relational database.

7. The computer implemented method of claim **4**, wherein the database is a multidimensional database.

8. The computer implemented method of claim **4**, wherein the database contains semantic extractions of another database.

9. The computer implemented method of claim **1**, wherein the dataset is retrieved from a remote database.

10. The computer implemented method of claim **2**, wherein the rules are predetermined.

11. The computer implemented method of claim **2**, wherein the rules relate to user preference.

12. The computer implemented method of claim **2**, wherein the rules relate to user usage.

13. The computer implemented method of claim **1**, the selecting step further comprising:

computing the cardinalities of all the fields in the resulting view.

14. The computer implemented method of claim **2**, wherein the rules relate to the cardinalities of all the fields in the resulting view.

15. The computer implemented method of claim **1**, the selecting step further comprising:

computing the functional dependencies of all the fields in the resulting view, using an algorithm.

16. The computer implemented method of claim **2**, wherein the rules relate to the functional dependency of all the fields in the resulting view.

17. The computer implemented method of claim **1**, the displaying of the dataset step further comprising:

choosing a mark for the resulting view; and

rendering the dataset according to the mark.

18. The computer implemented method of claim **1**, wherein the receiving step is accomplished when the user drags and drops a representation of a field.

19. The computer implemented method of claim **1**, wherein the receiving step is accomplished when the user double clicks a representation of a field.

20. The computer implemented method of claim **1**, wherein the receiving step is accomplished when the user right clicks a representation of a field and selects an option to drop or add the field.

21. The computer implemented method of claim **1**, wherein the receiving step is accomplished when the user inputs text to designate the plurality of user selected fields and their order.

22. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a text table when the resulting view's first field is of the dimension data type.

23. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a bar chart when the resulting view's first field is of the

measure data type, the resulting view's second field is of the dimension data type, and the resulting view has no field of the date dimension data type.

24. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a line when the resulting view's first field is of the measure data type, the resulting view's second field is of the dimension data type, and the resulting view has at least one field of the date dimension data type.

25. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a discrete line when the resulting view's first field is of the date dimension data type, and the resulting view's second field is of the measure data type.

26. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a continuous line when the resulting view's first field is of the continuous dimension data type, and the resulting view's second field is of the measure data type.

27. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a scatter plot when the resulting view's first and second fields are of the measure data type, and the resulting view's subsequent fields are of the dimension data type.

28. The computer implemented method of claim **2**, wherein the one or more rules state that the resulting view is a scatter matrix when the resulting view's first, second, and a subsequent field is of the measure data type.

29. The computer implemented method of claim **17**, wherein when the resulting view is a text table, the mark is text.

30. The computer implemented method of claim **17**, wherein when the resulting view is a heat map, the mark is a square.

31. The computer implemented method of claim **17**, wherein when the resulting view is a bar chart, the mark is a bar.

32. The computer implemented method of claim **17**, wherein when the resulting view is a line graph, the mark is a line segment.

33. The computer implemented method of claim **17**, wherein when the resulting view is a circle chart, the mark is a circle.

34. The computer implemented method of claim **17**, wherein when the resulting view is a Gantt plot, the mark is a bar.

35. The computer implemented method of claim **17**, wherein when the resulting view is a scatter plot, the mark is a point or shape.

36. The computer implemented method of claim **1**, wherein the selecting step uses heuristics.

37. A computer implemented method for displaying a graphical representation of a dataset with a plurality of tuples, comprising:

at a computer having one or more processors and memory storing programs executed by the one or more processors:

displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

receiving a user instruction for generating a graphical representation of the dataset, wherein the user instruction includes a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, each of the user-selected terms having an associated data type; and

in response to the user selection and the user ordering,

forming a plurality of alternative views based on the user ordering of the user-selected terms, each alternative view showing all tuples, or a transformation of all tuples, in the dataset;

determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system;

selecting a resulting view from the plurality of rated alternative views, based upon a user selected option and comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

displaying the dataset or a transformation of the dataset in the data visualization region according to the resulting view.

38. The computer implemented method of claim **37**, wherein the dataset is a database.

39. The computer implemented method of claim **38**, wherein the database is remote.

40. The computer implemented method of claim **38**, wherein the database is a relational database.

41. The computer implemented method of claim **38**, wherein the database is a multidimensional database.

42. The computer implemented method of claim **38**, wherein the database contains semantic extractions of another database.

43. The computer implemented method of claim **37**, wherein the dataset is retrieved from a remote database.

44. The computer implemented method of claim **37**, wherein the forming step further comprises:

computing the cardinality of the fields in the plurality of tuples.

45. The computer implemented method of claim **37**, wherein the forming step further comprises:

computing the functional dependency on the plurality of tuples, using an algorithm.

46. The computer implemented method of claim **37**, wherein, when the user selected option is a first option, the selecting step further comprises:

ranking the plurality of alternative views according to a rating system; and

assigning the resulting view as the highest ranked alternative view.

47. The computer implemented method of claim **46**, wherein the rating system is predetermined.

48. The computer implemented method of claim **46**, wherein the rating system relates to user preference.

49. The computer implemented method of claim **46**, wherein the rating system relates to user usage.

50. The computer implemented method of claim **37**, wherein, when the user selected option is a second option, the selecting step further comprises:

displaying a list of the alternative views;

receiving the user's selection of an alternative view; and

assigning the resulting view as the alternative view selected by the user.

51. The computer implemented method of claim **37**, wherein the forming step further comprises:

determining the applicable view types according to the dataset's data types;

identifying the associated mark for each applicable view type; and

forming an alternative view for each applicable view type according to the associated mark.

52. The computer implemented method of claim **51**, wherein the determining step is accomplished using a set of criteria.

53. The computer implemented method of claim **52**, wherein the set of criteria is predetermined.

54. The computer implemented method of claim **52**, wherein the set of criteria relates to user preference.

55. The computer implemented method of claim **52**, wherein the set of criteria relates to user usage.

56. The computer implemented method of claim **52**, wherein an applicable view type is a text table and the associated mark for such an applicable view type is text, if all the data contained in the dataset are aggregated according to their fields, and no data type is independently quantitative.

57. The computer implemented method of claim **52**, wherein an applicable view type is a heat map and the associated mark for such an applicable view type is a colored shape, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data type is independently quantitative.

58. The computer implemented method of claim **52**, wherein an applicable view type is a Gantt plot and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, less than three data types are dependently quantitative, and at least one data type is either independently quantitative or a date.

59. The computer implemented method of claim **52**, wherein an applicable view type is a side-by-side bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, at least one data type is dependently quantitative, and no data type is independently quantitative.

60. The computer implemented method of claim **52**, wherein an applicable view type is a stacked bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, no more than two data types are ordinal, and no data types are independently quantitative.

61. The computer implemented method of claim **52**, wherein an applicable view type is a measure bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data types are independently quantitative.

62. The computer implemented method of claim **52**, wherein an applicable view type is a dimension line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, at least one data type is a date, and no data types are independently quantitative.

63. The computer implemented method of claim **52**, wherein an applicable view type is a measure line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, and at least one data type is either independently quantitative or a date.

64. The computer implemented method of claim **52**, wherein an applicable view type is a circle chart and the associated mark for such an applicable view type is a circle, if at least one data type is ordinal, at least one data type is dependently quantitative, and no data type can be aggregated or is independently quantitative.

65. The computer implemented method of claim **52**, wherein an applicable view type is a single scatter chart and the associated mark for such an applicable view type is a point, if the data contained in the dataset included between two to four dependently quantitative data types, and no independently quantitative data type.

66. The computer implemented method of claim **52**, wherein an applicable view type is a matrix scatter chart and the associated mark for such an applicable view type is a point, if the dataset included at least three dependently quantitative data types, and no independently quantitative data type.

67. The computer implemented method of claim **52**, wherein an applicable view type is a histogram and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, and no data types are independently quantitative.

68. The computer implemented method of claim **37**, wherein the selecting step uses heuristics.

69. A computer program product including a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising instructions that are executed to display a graphical representation of a dataset, the instructions comprising:

a module for displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

a field receiver for receiving a user selection of at least a subset of the multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, wherein each user-selected term has an associated data type;

a resulting view selector for, in response to the user selection and the user ordering,

forming a plurality of alternative views using the dataset based on the user ordering of the user-selected terms;

determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system;

selecting a resulting view from the plurality of alternative views for displaying the dataset based on comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

a dataset displayer for displaying the dataset or a transformation of the dataset according to the resulting view in the data visualization region.

70. The computer program product of claim **69**, wherein the resulting view selector further comprises:

with no further user instruction:

a rule chooser for choosing one or more rules that are associated with adding the user selected fields; and

a rule applier for applying the one or more rules to determine the resulting view's view type.

71. The computer program product of claim **70**, further comprising:

a view determiner for determining whether a first view exists;

wherein the one or more rules selected is also based upon the first view, if the first view is determined to exist.

72. The computer program product of claim **69**, wherein the dataset is a database.

73. The computer program product of claim **72**, wherein the database is remote.

74. The computer program product of claim **72**, wherein the database is a relational database.

75. The computer program product of claim **72**, wherein the database is a multidimensional database.

76. The computer program product of claim **72**, wherein the database contains semantic extractions of another database.

77. The computer program product of claim **69**, wherein the dataset is retrieved from a remote database.

78. The computer program product of claim **70**, wherein the rules are predetermined.

79. The computer program product of claim **70**, wherein the rules relate to user preference.

80. The computer program product of claim **70**, wherein the rules relate to user usage.

81. The computer program product of claim **69**, the resulting view selector further comprising:

a cardinality computer computing the cardinalities of all the fields in the resulting view.

82. The computer program product of claim **70**, wherein the rules relate to the cardinalities of all the fields in the resulting view.

83. The computer program product of claim **69**, the resulting view selector further comprising:

a functional dependency computer for computing the functional dependencies of all the fields in the resulting view, using an algorithm.

84. The computer program product of claim **70**, wherein the rules relate to the functional dependency of all the fields in the resulting view.

85. The computer program product of claim **69**, the dataset displayer further comprising:

a mark chooser for choosing a mark for the resulting view; and

a dataset renderer for rendering the dataset according to the mark.

86. The computer program product of claim **69**, wherein the field receiver allows the user to drag and drop a representation of a field.

87. The computer program product of claim **69**, wherein the field receiver allows the user to double click a representation of a field.

88. The computer program product of claim **69**, wherein the field receiver allows the user to right click a representation of a field and selects an option to drop or add the field.

89. The computer program product of claim **69**, wherein the field receiver allows the user to designate the plurality of user selected fields and their order.

90. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a text table when the resulting view's first field is of the dimension data type.

91. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a bar chart when the resulting view's first field is of the measure data type, the resulting view's second field is of the dimension data type, and the resulting view has no field of the date dimension data type.

92. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a line when the resulting view's first field is of the measure data type, the resulting view's second field is of the dimension data type, and the resulting view has at least one field of the date dimension data type.

93. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a discrete line when the resulting view's first field is of the date dimension data type, and the resulting view's second field is of the measure data type.

94. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a continuous line when the resulting view's first field is of the continuous dimension data type, and the resulting view's second field is of the measure data type.

95. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a scatter plot when the resulting view's first and second fields are of the measure data type, and the resulting view's subsequent fields are of the dimension data type.

96. The computer program product of claim **70**, wherein the one or more rules state that the resulting view is a scatter matrix when the resulting view's first, second, and a subsequent field is of the measure data type.

97. The computer program product of claim **85**, wherein when the resulting view is a text table, the mark is text.

98. The computer program product of claim **85**, wherein when the resulting view is a heat map, the mark is a square.

99. The computer program product of claim **85**, wherein when the resulting view is a bar chart, the mark is a bar.

100. The computer program product of claim **85**, wherein when the resulting view is a line graph, the mark is a line segment.

101. The computer program product of claim **85**, wherein when the resulting view is a circle chart, the mark is a circle.

102. The computer program product of claim **85**, wherein when the resulting view is a Gantt plot, the mark is a bar.

103. The computer program product of claim **85**, wherein when the resulting view is a scatter plot, the mark is a point or shape.

104. The computer program product of claim **69**, wherein the selecting step uses heuristics.

105. A computer program product including a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising instructions that are executed to display a graphical representation of a dataset with a plurality of tuples, the instructions comprising:

a module for displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

an alternative view former for forming a plurality of alternative views in response to a user instruction for generating a graphical representation of the dataset, each alternative view showing all tuples, or a transformation of all tuples, in the dataset, wherein the user instruction includes a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, each of the user-selected terms having an associated data type, and determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system, wherein the user instruction includes a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-

selected terms in the data visualization region, each of the user-selected terms having an associated data type;

a resulting view selector for selecting a resulting view from the plurality of rated alternative views in response to the user instruction, based upon a user selected option and comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

a dataset displayer for displaying the dataset or a transformation of the dataset in the data visualization region according to the resulting view in response to the user instruction.

106. The computer program product of claim **105**, wherein the dataset is a database.

107. The computer program product of claim **106**, wherein the database is remote.

108. The computer program product of claim **106**, wherein the database is a relational database.

109. The computer program product of claim **106**, wherein the database is a multidimensional database.

110. The computer program product of claim **106**, wherein the database contains semantic extractions of another database.

111. The computer program product of claim **105**, wherein the dataset is retrieved from a remote database.

112. The computer program product of claim **105**, wherein the alternative view former further comprises:

a cardinality computer for computing the cardinality of the fields in the plurality of tuples.

113. The computer program product of claim **105**, wherein the alternative view former further comprises:

a functional dependency computer for computing the functional dependency on the plurality of tuples, using an algorithm.

114. The computer program product of claim **105**, wherein, when the user selected option is a first option, the resulting view selector further comprises:

an alternative view ranker for ranking the plurality of alternative views according to a rating system; and

a view assignor for assigning the resulting view as the highest ranked alternative view.

115. The computer program product of claim **114**, wherein the rating system is predetermined.

116. The computer program product of claim **114**, wherein the rating system relates to user preference.

117. The computer program product of claim **114**, wherein the rating system relates to user usage.

118. The computer program product of claim **105**, wherein, when the user selected option is a second option, the resulting view selector further comprises:

a list displayer for displaying a list of the alternative views;

a selection receiver for receiving the user's selection of an alternative view; and

a view assignor for assigning the resulting view as the alternative view selected by the user.

119. The computer program product of claim **105**, wherein the alternative view former forms an alternative view for each applicable view type according to an associated mark, the alternative view former further comprising:

a view type determiner for determining the applicable view types according to the dataset's data types; and

a mark chooser for choosing the associated mark for each applicable view type.

120. The computer program product of claim **119**, wherein the view type determiner uses a set of criteria.

121. The computer program product of claim **120**, wherein the set of criteria is predetermined.

122. The computer program product of claim **120**, wherein the set of criteria relates to user preference.

123. The computer program product of claim **120**, wherein the set of criteria relates to user usage.

124. The computer program product of claim **120**, wherein an applicable view type is a text table and the associated mark for such an applicable view type is text, if all the data contained in the dataset are aggregated according to their fields, and no data type is independently quantitative.

125. The computer program product of claim **120**, wherein an applicable view type is a heat map and the associated mark for such an applicable view type is a colored shape, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data type is independently quantitative.

126. The computer program product of claim **120**, wherein an applicable view type is a Gantt plot and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, less than three data types are dependently quantitative, and at least one data type is either independently quantitative or a date.

127. The computer program product of claim **120**, wherein an applicable view type is a side-by-side bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, at least one data type is dependently quantitative, and no data type is independently quantitative.

128. The computer program product of claim **120**, wherein an applicable view type is a stacked bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, no more than two data types are ordinal, and no data types are independently quantitative.

129. The computer program product of claim **120**, wherein an applicable view type is a measure bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data types are independently quantitative.

130. The computer program product of claim **120**, wherein an applicable view type is a dimension line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, at least one data type is a date, and no data types are independently quantitative.

131. The computer program product of claim **120**, wherein an applicable view type is a measure line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, and at least one data type is either independently quantitative or a date.

132. The computer program product of claim **120**, wherein an applicable view type is a circle chart and the associated mark for such an applicable view type is a circle, if at least one data type is ordinal, at least one data type is dependently quantitative, and no data type can be aggregated or is independently quantitative.

133. A computer system for automatically and visually displaying a graphical representation of a dataset, the computer system comprising:

a central processing unit;

a memory, coupled to the central processing unit, the memory storing:

the dataset;

a programming module comprising, comprising:

instructions for displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

instructions for receiving a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, wherein each field has an associated data type;

instructions for forming a plurality of alternative views using the dataset based on the user ordering of the user-selected terms;

instructions for determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system and selecting a resulting view from the plurality of alternative views for displaying the dataset based on comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

instructions for displaying the dataset or a transformation of the dataset in the data visualization region according to the resulting view.

134. The computer system of claim **133**, wherein the instructions for selecting further comprises:

with no further user instruction:

instructions for choosing one or more rules that are associated with adding the user selected fields; and

instructions for applying the one or more rules to determine the resulting view's view type.

135. The computer system of claim **134**, further comprising, before the instructions for choosing:

instructions for determining whether a first view exists;

wherein the one or more rules selected is also based upon the first view, if the first view is determined to exist.

136. The computer system of claim **133**, wherein the dataset is a database.

137. The computer system of claim **136**, wherein the database is remote.

138. The computer system of claim **136**, wherein the database is a relational database.

139. The computer system of claim **136**, wherein the database is a multidimensional database.

140. The computer system of claim **136**, wherein the database contains semantic extractions of another database.

141. The computer system of claim **133**, wherein the dataset is retrieved from a remote database.

142. The computer system of claim **134**, wherein the rules are predetermined.

143. The computer system of claim **134**, wherein the rules relate to user preference.

144. The computer system of claim **134**, wherein the rules relate to user usage.

145. The computer system of claim **133**, the instructions for selecting further comprising:

instructions for computing the cardinalities of all the fields in the resulting view.

146. The computer system of claim **134**, wherein the rules relate to the cardinalities of all the fields in the resulting view.

147. The computer system of claim **133**, the instructions for selecting further comprising:

instructions for computing the functional dependencies of all the fields in the resulting view, using an algorithm.

148. The computer system of claim **134**, wherein the rules relate to the functional dependency of all the fields in the resulting view.

149. The computer system of claim **133**, the instructions for displaying the dataset further comprising:

instructions for choosing a mark for the resulting view; and

instructions for rendering the dataset according to the mark.

150. The computer system of claim **133**, wherein the instructions for receiving is accomplished when the user drags and drops a representation of a field.

151. The computer system of claim **133**, wherein the instructions for receiving is accomplished when the user double clicks a representation of a field.

152. The computer system of claim **133**, wherein the instructions for receiving is accomplished when the user right clicks a representation of a field and selects an option to drop or add the field.

153. The computer system of claim **133**, wherein the instructions for receiving is accomplished when the user inputs text to designate the plurality of user selected fields and their order.

154. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a text table when the resulting view's first field is of the dimension data type.

155. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a bar chart when the resulting view's first field is of the measure data type, the resulting view's second field is of the dimension data type, and the resulting view has no field of the date dimension data type.

156. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a line when the resulting view's first field is of the measure data type, the resulting view's second field is of the dimension data type, and the resulting view has at least one field of the date dimension data type.

157. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a discrete line when the resulting view's first field is of the date dimension data type, and the resulting view's second field is of the measure data type.

158. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a continuous line when the resulting view's first field is of the continuous dimension data type, and the resulting view's second field is of the measure data type.

159. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a scatter plot when the resulting view's first and second fields are of the measure data type, and the resulting view's subsequent fields are of the dimension data type.

160. The computer system of claim **134**, wherein the one or more rules state that the resulting view is a scatter matrix when the resulting view's first, second, and a subsequent field is of the measure data type.

161. The computer system of claim **149**, wherein when the resulting view is a text table, the mark is text.

162. The computer system of claim **149**, wherein when the resulting view is a heat map, the mark is a square.

163. The computer system of claim **149**, wherein when the resulting view is a bar chart, the mark is a bar.

164. The computer system of claim **149**, wherein when the resulting view is a line graph, the mark is a line segment.

165. The computer system of claim **149**, wherein when the resulting view is a circle chart, the mark is a circle.

166. The computer system of claim **149**, wherein when the resulting view is a Gantt plot, the mark is a bar.

167. The computer system of claim **149**, wherein when the resulting view is a scatter plot, the mark is a point or shape.

168. The computer system of claim **133**, wherein the selecting step uses heuristics.

169. A computer system for automatically and visually displaying a graphical representation of a dataset with a plurality of tuples, the computer system comprising:

a central processing unit;

a memory, coupled to the central processing unit, the memory storing:

the dataset;

a programming module comprising:

instructions for displaying a graphical user interface window including a schema display region and a data visualization region, wherein the schema display region includes multiple terms corresponding to a plurality of fields of the dataset, wherein each field defines a respective category of the dataset;

instructions for forming a plurality of alternative views in response to a user instruction for generating a graphical representation of the dataset, each alternative view showing all tuples, or a transformation of all tuples, in the dataset, wherein the user instruction includes a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, each of the user-selected terms having an associated data type, and determining a rank for each of the plurality of alternative views based on the user-specified order and the associated data types of the user-selected terms and a rating system, wherein the user instruction includes a user selection of at least a subset of multiple terms in the schema display region and a user ordering of the user-selected terms in the data visualization region, each of the user-selected terms having an associated data type;

instructions for selecting a resulting view from the plurality of rated alternative views in response to the user instruction, based upon a user selected option and comparing the rank of the resulting view with the rank of at least another view of the plurality of alternative views; and

instructions for displaying the dataset or a transformation of the dataset in the data visualization region according to the resulting view in response to the user instruction.

170. The computer system of claim **169**, wherein the dataset is a database.

171. The computer system of claim **170**, wherein the database is remote.

172. The computer system of claim **170**, wherein the database is a relational database.

173. The computer system of claim **170**, wherein the database is a multidimensional database.

174. The computer system of claim **170**, wherein the database contains semantic extractions of another database.

175. The computer system of claim **169**, wherein the dataset is retrieved from a remote database.

176. The computer system of claim **169**, wherein the instructions for forming further comprises:

instructions for computing the cardinality of the fields in the plurality of tuples.

177. The computer system of claim **169**, wherein the instructions for forming further comprises:

instructions for computing the functional dependency on the plurality of tuples, using an algorithm.

178. The computer system of claim **169**, wherein, when the user selected option is a first option, the instructions for selecting further comprises:

instructions for ranking the plurality of alternative views according to a rating system; and

instructions for assigning the resulting view as the highest ranked alternative view.

179. The computer system of claim **178**, wherein the rating system is predetermined.

180. The computer system of claim **178**, wherein the rating system relates to user preference.

181. The computer system of claim **178**, wherein the rating system relates to user usage.

182. The computer system of claim **169**, wherein, when the user selected option is a second option, the instructions for selecting further comprises:

instructions for displaying a list of the alternative views;

instructions for receiving the user's selection of an alternative view; and

instructions for assigning the resulting view as the alternative view selected by the user.

183. The computer system of claim **169**, wherein the instructions for forming further comprises:

instructions for determining the applicable view types according to the dataset's data types;

instructions for identifying the associated mark for each applicable view type; and

instructions for forming an alternative view for each applicable view type according to the associated mark.

184. The computer system of claim **183**, wherein the instructions for determining is accomplished using a set of criteria.

185. The computer system of claim **184**, wherein the set of criteria is predetermined.

186. The computer system of claim **184**, wherein the set of criteria relates to user preference.

187. The computer system of claim **184**, wherein the set of criteria relates to user usage.

188. The computer system of claim **184**, wherein an applicable view type is a text table and the associated mark for such an applicable view type is text, if all the data contained in the dataset are aggregated according to their fields, and no data type is independently quantitative.

189. The computer system of claim **184**, wherein an applicable view type is a heat map and the associated mark for such an applicable view type is a colored shape, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data type is independently quantitative.

190. The computer system of claim **184**, wherein an applicable view type is a Gantt plot and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is ordinal, less than three data types are dependently quantitative, and at least one data type is either independently quantitative or a date.

191. The computer system of claim **184**, wherein an applicable view type is a side-by-side bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their

fields, at least one data type is ordinal, at least one data type is dependently quantitative, and no data type is independently quantitative.

192. The computer system of claim **184**, wherein an applicable view type is a stacked bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, no more than two data types are ordinal, and no data types are independently quantitative.

193. The computer system of claim **184**, wherein an applicable view type is a measure bar chart and the associated mark for such an applicable view type is a bar, if all the data contained in dataset are aggregated according to their fields, at least one data type is ordinal, no more than two data types are dependently quantitative, and no data types are independently quantitative.

194. The computer system of claim **184**, wherein an applicable view type is a dimension line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, at least one data type is a date, and no data types are independently quantitative.

195. The computer system of claim **184**, wherein an applicable view type is a measure line graph and the associated mark for such an applicable view type is a line, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, and at least one data type is either independently quantitative or a date.

196. The computer system of claim **184**, wherein an applicable view type is a circle chart and the associated mark for such an applicable view type is a circle, if at least one data type is ordinal, at least one data type is dependently quantitative, and no data type can be aggregated or is independently quantitative.

197. The computer system of claim **184**, wherein an applicable view type is a single scatter chart and the associated mark for such an applicable view type is a point, if the data contained in the dataset included between two to four dependently quantitative data types, and no independently quantitative data type.

198. The computer system of claim **184**, wherein an applicable view type is a matrix scatter chart and the associated mark for such an applicable view type is a point, if the dataset included at least three dependently quantitative data types, and no independently quantitative data type.

199. The computer system of claim **184**, wherein an applicable view type is a histogram and the associated mark for such an applicable view type is a bar, if all the data contained in the dataset are aggregated according to their fields, at least one data type is dependently quantitative, and no data types are independently quantitative.

200. The computer system of claim **169**, wherein the selecting step uses heuristics.

201. The computer implemented method of claim **1**, the selecting step further comprising:

displaying the plurality of alternative views; and

receiving a user selection of one of the plurality of alternative views as the resulting view.

202. The computer implemented method of claim **37**, the selecting step further comprising:

displaying the plurality of alternative views; and

receiving a user selection of one of the plurality of alternative views as the resulting view.

* * * * *