Visualization Analysis & Design

Data Abstraction (Ch 2)

Tamara Munzner
Department of Computer Science
University of British Columbia
@tamaramunzner

What does data mean?

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Basil, 7, S, Pear

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Basil, 7, S, Pear

- What about this data?

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Basil, 7, S, Pear

- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Basil, 7, S, Pear

- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
  - Basil Point neighborhood of city had 7 inches of snow cleared by the Pear Creek Limited snow removal service

What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

Basil, 7, S, Pear

- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
  - Basil Point neighborhood of city had 7 inches of snow cleared by the Pear Creek Limited snow removal service
  - lab rat Basil made 7 attempts to find way through south section of maze, these trials used pear as reward food

Now what?

- semantics: real-world meaning

Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Now what?

- semantics: real-world meaning

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Now what?

- semantics: real-world meaning
- data types: structural or mathematical interpretation of data
  - item, link, attribute, position, (grid)
  - different from data types in programming!

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

item: person

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

item: person

Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

attributes: name, age, shirt size, fave fruit

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

item: person

Other data types

- links
  - express relationship between two items
  - eg friendship on facebook, interaction between proteins
- positions
  - spatial data: location in 2D or 3D
  - pixels in photo, voxels in MRI scan, latitude/longitude
- grids
  - sampling strategy for continuous data

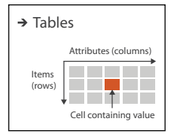
Dataset types

Tables

Items

Attributes

- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair



attributes: name, age, shirt size, fave fruit

Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

item: person

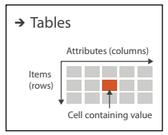
Dataset types

Tables

Items

Attributes

- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair
  - unique key (could be implicit)



attributes: name, age, shirt size, fave fruit

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

item: person

Table

	A	B	C	S	T	U
	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
	3	10/14/06	5-Low	Large Box	0.8	10/21/06
	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
	32	7/16/07	2-High	Small Pack	0.79	7/17/07
	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
	32	7/16/07	2-High	Medium Box	0.6	7/18/07
	32	7/16/07	2-High	Medium Box	0.65	7/18/07
	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
	70	12/18/06	5-Low	Small Box	0.59	12/23/06
	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
	96	4/17/05	2-High	Small Box	0.55	4/19/05
	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
	129	11/19/08	5-Low	Small Box	0.37	11/28/08
	130	5/8/08	2-High	Small Box	0.37	5/9/08
	130	5/8/08	2-High	Medium Box	0.38	5/10/08
	130	5/8/08	2-High	Small Box	0.6	5/11/08
	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
	166	9/12/07	2-High	Small Box	0.55	9/14/07
	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

item

	A	B	C	S	T	U
	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
	3	10/14/06	5-Low	Large Box	0.8	10/21/06
	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
	32	7/16/07	2-High	Small Pack	0.79	7/17/07
	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
	32	7/16/07	2-High	Medium Box	0.6	7/18/07
	32	7/16/07	2-High	Medium Box	0.65	7/18/07
	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
	70	12/18/06	5-Low	Small Box	0.59	12/23/06
	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
	96	4/17/05	2-High	Small Box	0.55	4/19/05
	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
	129	11/19/08	5-Low	Small Box	0.37	11/28/08
	130	5/8/08	2-High	Small Box	0.37	5/9/08
	130	5/8/08	2-High	Medium Box	0.38	5/10/08
	130	5/8/08	2-High	Small Box	0.6	5/11/08
	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
	166	9/12/07	2-High	Small Box	0.55	9/14/07
	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

item

attribute

	A	B	C	S	T	U
	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
	3	10/14/06	5-Low	Large Box	0.8	10/21/06
	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
	32	7/16/07	2-High	Small Pack	0.79	7/17/07
	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
	32	7/16/07	2-High	Medium Box	0.6	7/18/07
	32	7/16/07	2-High	Medium Box	0.65	7/18/07
	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
	70	12/18/06	5-Low	Small Box	0.59	12/23/06
	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
	96	4/17/05	2-High	Small Box	0.55	4/19/05
	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
	129	11/19/08	5-Low	Small Box	0.37	11/28/08
	130	5/8/08	2-High	Small Box	0.37	5/9/08
	130	5/8/08	2-High	Medium Box	0.38	5/10/08
	130	5/8/08	2-High	Small Box	0.6	5/11/08
	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
	166	9/12/07	2-High	Small Box	0.55	9/14/07
	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Table

item

attribute

	A	B	C	S	T	U
	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
	3	10/14/06	5-Low	Large Box	0.8	10/21/06
	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
	32	7/16/07	2-High	Small Pack	0.79	7/17/07
	32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
	32	7/16/07	2-High	Medium Box	0.6	7/18/07
	32	7/16/07	2-High	Medium Box	0.65	7/18/07
	35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
	35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
	36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
	65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
	66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
	69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
	69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
	70	12/18/06	5-Low	Small Box	0.59	12/23/06
	70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
	96	4/17/05	2-High	Small Box	0.55	4/19/05
	97	1/29/06	3-Medium	Small Box	0.38	1/30/06
	129	11/19/08	5-Low	Small Box	0.37	11/28/08
	130	5/8/08	2-High	Small Box	0.37	5/9/08
	130	5/8/08	2-High	Medium Box	0.38	5/10/08
	130	5/8/08	2-High	Small Box	0.6	5/11/08
	132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
	132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
	134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
	135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
	166	9/12/07	2-High	Small Box	0.55	9/14/07
	193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
	194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

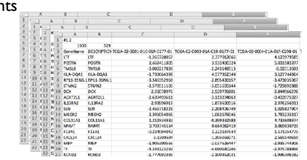
Dataset types

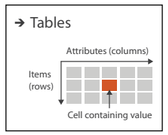
Tables

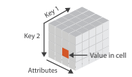
Items

Attributes

- multidimensional tables
  - indexing based on multiple keys
    - eg genes, patients







Dataset types

Tables

Items

Attributes

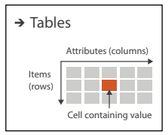
Networks & Trees

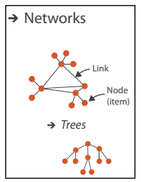
Items (nodes)

Links

Attributes

- network/graph
  - nodes (vertices) connected by links (edges)
  - tree is special case: no cycles
    - often have roots and are directed





Dataset types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

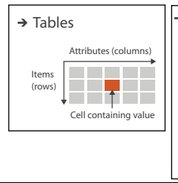
Attributes

Fields

Grids

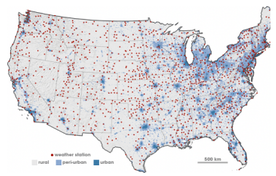
Positions

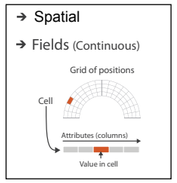
Attributes



Spatial fields

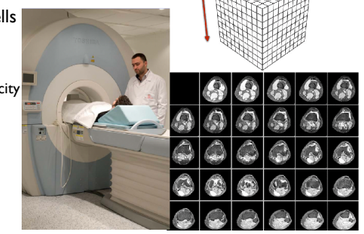
- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated





Spatial fields

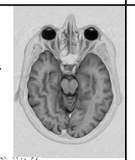
- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling: where attributes are measured
  - interpolation: how to model attributes elsewhere
  - grid types



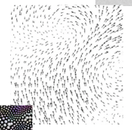
Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling: where attributes are measured
  - interpolation: how to model attributes elsewhere
  - grid types
- major divisions
  - attributes per cell: scalar (1), vector (2), tensor (many)

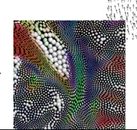
scalar



vector



tensor



Dataset types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

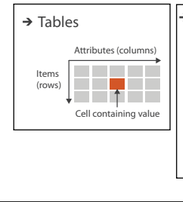
Positions

Attributes

Geometry

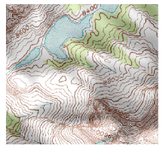
Items

Positions



Geometry

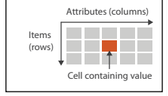
- shape of items
- explicit spatial positions / regions
  - points, lines, curves, surfaces, volumes
- boundary between computer graphics and visualization
  - graphics: geometry taken as given
  - vis: geometry is result of a design decision



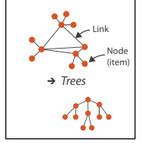
Dataset types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

→ Tables

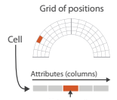


→ Networks



→ Spatial

→ Fields (Continuous)



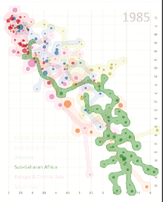
→ Geometry (Spatial)

Collections

- how we group items

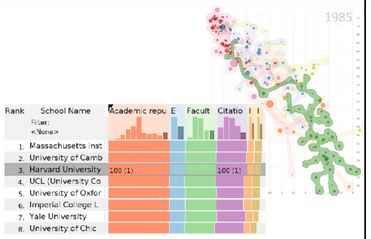
Collections

- how we group items
- sets
  - unique items, unordered



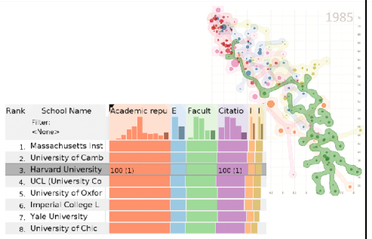
Collections

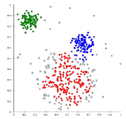
- how we group items
- sets
  - unique items, unordered
- lists
  - ordered, duplicates possible



Collections

- how we group items
- sets
  - unique items, unordered
- lists
  - ordered, duplicates possible
- clusters
  - groups of similar items





Dataset and data types

→ Data and Dataset Types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

→ Data Types

→ Items → Attributes → Links → Positions → Grids

Attribute types

- which classes of values & measurements?
- categorical (nominal)
  - compare equality
  - no implicit ordering
- ordered
  - ordinal
    - less/greater than defined
  - quantitative
    - meaningful magnitude
    - arithmetic possible

→ Attribute Types

→ Categorical → Ordered

→ Ordinal → Quantitative

Table

Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

categorical
ordinal
quantitative

Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06

Other data concerns

→ Attribute Types

→ Categorical → Ordered

→ Ordinal → Quantitative

→ Ordering Direction

→ Sequential → Diverging → Cyclic

→ Static → Dynamic

→ Dataset Availability

Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
  - how many items in the dataset?
  - what is cardinality of each attribute?
    - number of levels for categorical data
    - range for quantitative data
- consider whether to transform data
  - guided by understanding of task

Data vs conceptual models

- data model
  - mathematical abstraction
    - sets with operations, eg floats with * / - +
    - variable data types in programming languages
- conceptual model
  - mental construction (semantics)
  - supports reasoning
  - typically based on understanding of tasks [stay tuned!]
- data abstraction process relies on conceptual model
  - for transforming data if needed

Data vs conceptual model, example

Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...

Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature

Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions

Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather

Data vs conceptual model, example

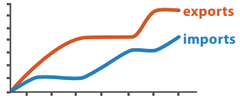
- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather
  - hot, warm, cold: ordinal
    - task: deciding if bath water is ready

Data vs conceptual model, example

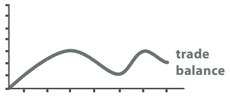
- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather
  - hot, warm, cold: ordinal
    - task: deciding if bath water is ready
  - above freezing, below freezing: categorical
    - task: decide if I should leave the house today

Derived attributes

- derived attribute: compute from originals
  - simple change of type
  - acquire additional data
  - complex transformation



Original Data



trade balance = exports – imports

Derived Data

Analysis example: Derive one attribute

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton

[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]