

Data Abstraction - Summary

- 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

Dataset types

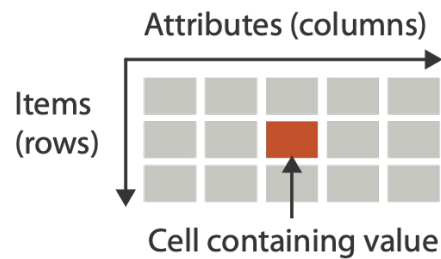
Tables

Items

Attributes

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

→ Tables



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

Dataset types

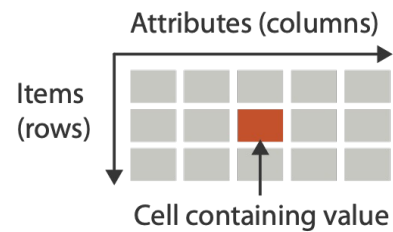
Tables

Items

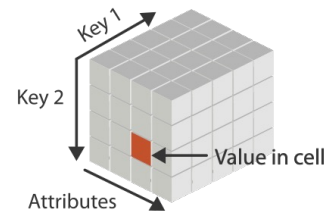
Attributes

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

→ Tables

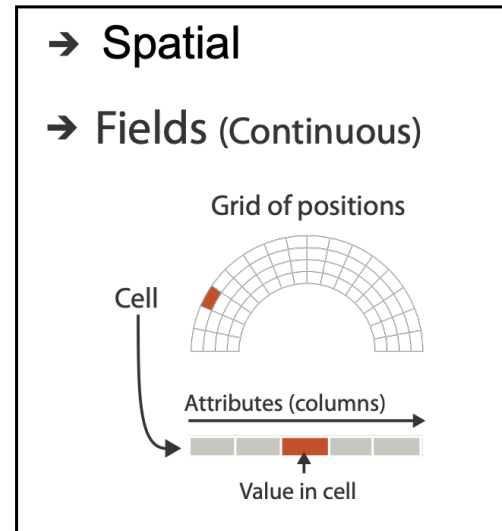
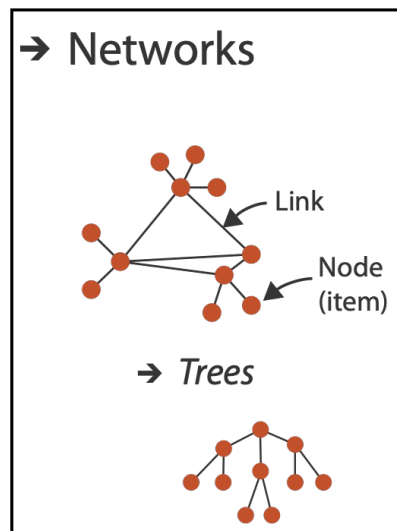
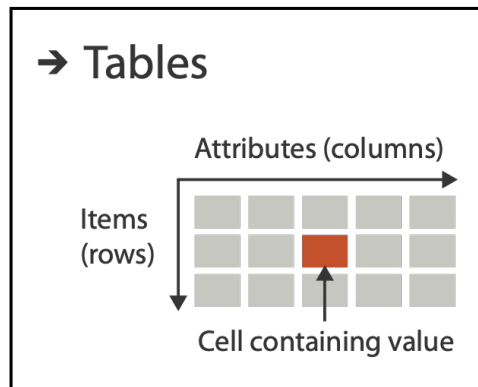
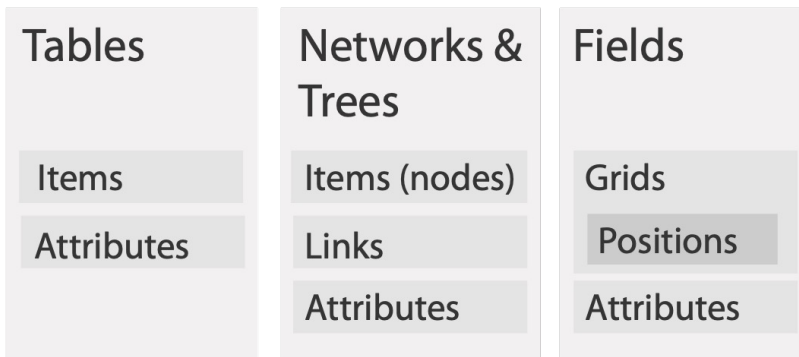


→ Multidimensional Table



1	A	B	C	D	E
2	A	B	C	D	E
3	1	1	1	1	1
4	2	2	2	2	2
5	3	3	3	3	3
6	4	4	4	4	4
7	5	5	5	5	5
8	6	6	6	6	6
9	7	7	7	7	7
10	8	8	8	8	8
11	9	9	9	9	9
12	10	10	10	10	10
13	11	11	11	11	11
14	12	12	12	12	12
15	13	13	13	13	13
16	14	14	14	14	14
17	15	15	15	15	15
18	16	16	16	16	16
19	17	17	17	17	17
20	18	18	18	18	18
21	19	19	19	19	19
22	20	20	20	20	20
23	21	21	21	21	21
24	22	22	22	22	22
25	23	23	23	23	23
26	24	24	24	24	24
27	25	25	25	25	25
28	26	26	26	26	26
29	27	27	27	27	27
30	28	28	28	28	28
31	29	29	29	29	29
32	30	30	30	30	30
33	31	31	31	31	31
34	32	32	32	32	32
35	33	33	33	33	33
36	34	34	34	34	34
37	35	35	35	35	35
38	36	36	36	36	36
39	37	37	37	37	37
40	38	38	38	38	38
41	39	39	39	39	39
42	40	40	40	40	40
43	41	41	41	41	41
44	42	42	42	42	42
45	43	43	43	43	43
46	44	44	44	44	44
47	45	45	45	45	45
48	46	46	46	46	46
49	47	47	47	47	47
50	48	48	48	48	48
51	49	49	49	49	49
52	50	50	50	50	50
53	51	51	51	51	51
54	52	52	52	52	52
55	53	53	53	53	53
56	54	54	54	54	54
57	55	55	55	55	55
58	56	56	56	56	56
59	57	57	57	57	57
60	58	58	58	58	58
61	59	59	59	59	59
62	60	60	60	60	60
63	61	61	61	61	61
64	62	62	62	62	62
65	63	63	63	63	63
66	64	64	64	64	64
67	65	65	65	65	65
68	66	66	66	66	66
69	67	67	67	67	67
70	68	68	68	68	68
71	69	69	69	69	69
72	70	70	70	70	70
73	71	71	71	71	71
74	72	72	72	72	72
75	73	73	73	73	73
76	74	74	74	74	74
77	75	75	75	75	75
78	76	76	76	76	76
79	77	77	77	77	77
80	78	78	78	78	78
81	79	79	79	79	79
82	80	80	80	80	80
83	81	81	81	81	81
84	82	82	82	82	82
85	83	83	83	83	83
86	84	84	84	84	84
87	85	85	85	85	85
88	86	86	86	86	86
89	87	87	87	87	87
90	88	88	88	88	88
91	89	89	89	89	89
92	90	90	90	90	90
93	91	91	91	91	91
94	92	92	92	92	92
95	93	93	93	93	93
96	94	94	94	94	94
97	95	95	95	95	95
98	96	96	96	96	96
99	97	97	97	97	97
100	98	98	98	98	98
101	99	99	99	99	99
102	100	100	100	100	100
103	101	101	101	101	101
104	102	102	102	102	102
105	103	103	103	103	103
106	104	104	104	104	104
107	105	105	105	105	105
108	106	106	106	106	106
109	107	107	107	107	107
110	108	108	108	108	108
111	109	109	109	109	109
112	110	110	110	110	110
113	111	111	111	111	111
114	112	112	112	112	112
115	113	113	113	113	113
116	114	114	114	114	114
117	115	115	115	115	115
118	116	116	116	116	116
119	117	117	117	117	117
120	118	118	118	118	118
121	119	119	119	119	119
122	120	120	120	120	120
123	121	121	121	121	121
124	122	122	122	122	122
125	123	123	123	123	123
126	124	124	124	124	124
127	125	125	125	125	125
128	126	126	126	126	126
129	127	127	127	127	127
130	128	128	128	128	128
131	129	129	129	129	129
132	130	130	130	130	130
133	131	131	131	131	131
134	132	132	132	132	132
135	133	133	133	133	133
136	134	134	134	134	134
137	135	135	135	135	135
138	136	136	136	136	136
139	137	137	137	137	137
140	138	138	138	138	138
141	139	139	139	139	139
142	140	140	140	140	140
143	141	141	141	141	141
144	142	142	142	142	142
145	143	143	143	143	143
146	144	144	144	144	144
147	145	145	145	145	145
148	146	146	146	146	146
149	147	147	147	147	147
150	148	148	148	148	148
151	149	149	149	149	149
152	150	150	150	150	150
153	151	151	151	151	151
154	152	152	152	152	152
155	153	153	153	153	153
156	154	154	154	154	154
157	155	155	155	155	155
158	156	156	156	156	156
159	157	157	157	157	157
160	158	158	158	158	158
161	159	159	159	159	159
162	160	160	160	160	160
163	161	161	161	161	161
164	162	162	162	162	162
165	163	163	163	163	163
166	164	164	164	164	164
167	165	165	165	165	165
168	166	166	166	166	166
169	167	167	167	167	167
170	168	168	168	168	168
171	169	169	169	169	169
172	170	170	170	170	170
173	171	171	171	171	171
174	172	172	172	172	172
175	173	173	173	173	173
176	174	174	174	174	174
177	175	175	175	175	175
178	176	176	176	176	176
179	177	177	177	177	177
180	178	178	178	178	178
181	179	179	179	179	179
182	180	180	180	180	180
183	181	181	181	181	181
184	182	182	182	182	182
185	183	183	183	183	183
186	184	184	184	184	184
187	185	185	185	185	185
188	186	186	186	186	186
189	187	187	187	187	187
190	188	188	188	188	188
191	189	189	189	189	189
192	190	190	190	190	190
193	191	191	191	191	191
194	192	192	192	192	192
195	193	193	193	193	193
196	194	194	194	194	194
197	195	195	195	195	195
198	196	196	196	196	196
199	197	197	197	197	197
200	198	198	198	198	198
201	199	199	199	199	199
202	200	200	200	200	200
203	201	201	201	201	201
204	202	202	202	202	202
205	203	203	203	203	203
206	204	204	204	204	204
207	205	205	205	205	205
208	206	206	206	206	206
209	207	207	207	207	207
210	208	208	208	208	208
211	209	209	209	209	209
212	210	210	210	210	210
213	211	211	211	211	211
214	212	212	212	212	212
215	213	213	213	213	213
216	214	214	214	214	214
217	215	215	215	215	215
218	216	216	216	216	216
219	217	217	217	217	217
220	218	218	218	218	218
221	219	219	219	219	219
222	220	220	220	220	220
223	221	221	221	221	221
224	222	222	222	222	222
225	223	223	223	223	223
226	224	224	224	224	224
227	225	225	225	225	225
228	226	226	226	226	226
229	227	227	227	227	227
230	228	228	228	228	228
231	229	229	229	229	229
232	230	230	230	230	230
233	231	231	231	231	231
234	232	232	232	232	232
235	233	233	233	233	233
236	234	234	234	234	234
237	235	235	235	235	235
238	236	236	236	236	236
239	237	237	237	237	237
240	238	238	238	238	238
241	239	239	239	239	239
242	240	240	240	240	240
243	241	241	241	241	241
244	242	242	242	242	242
245	243	243	243	243	243
246	244	244	244	244	244
247	245	245	245	245	245
248	246	246	246	246	246
249					

Dataset types

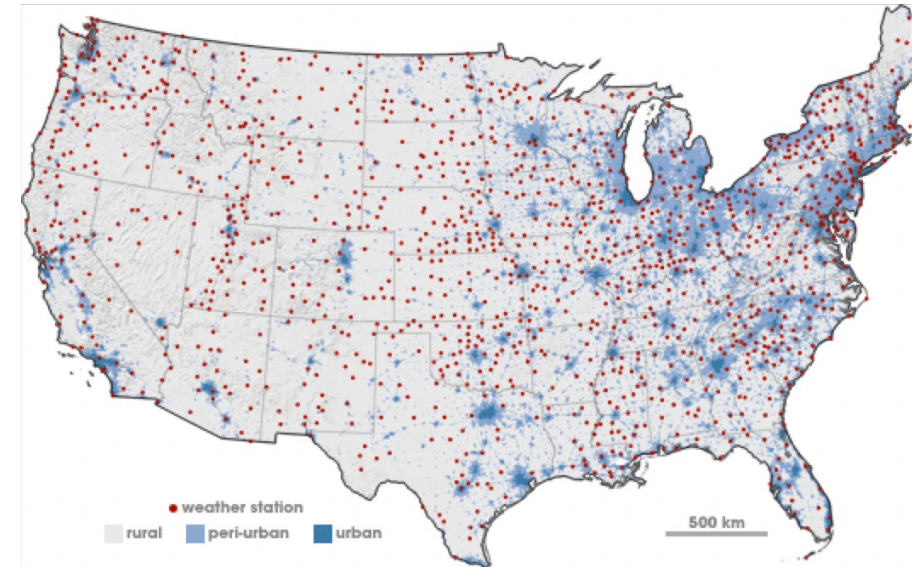
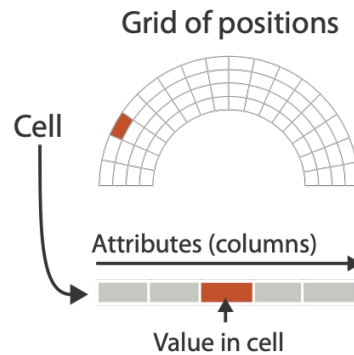


Spatial fields

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

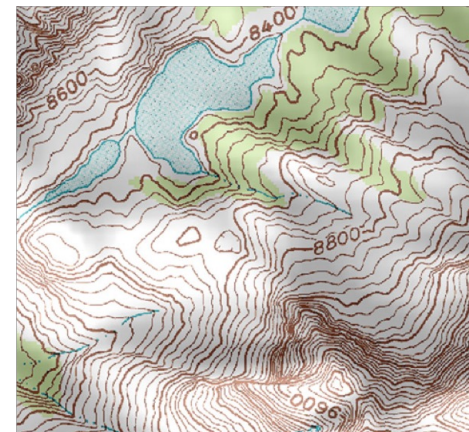
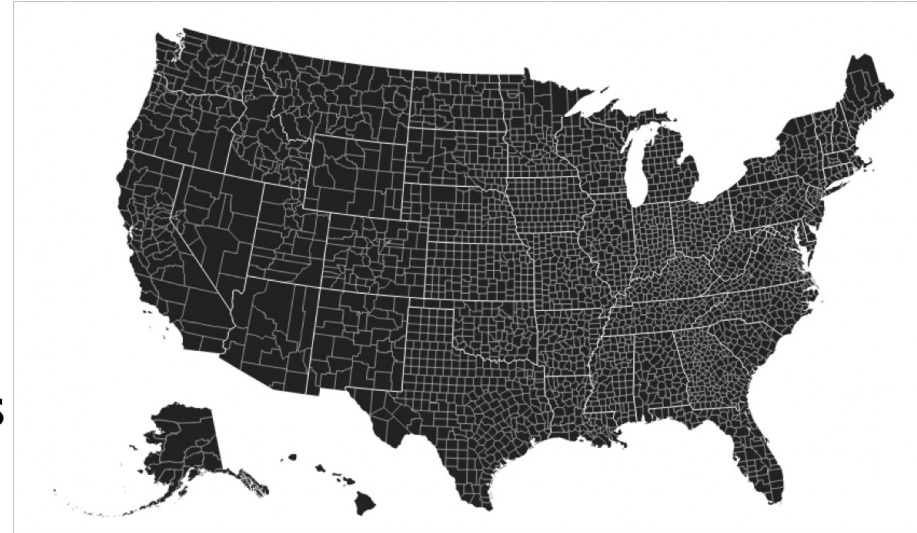
→ Spatial

→ Fields (Continuous)



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

Items

Attributes

Networks &
Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

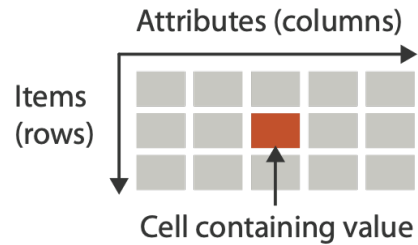
Attributes

Geometry

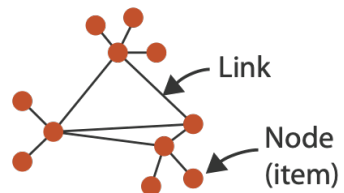
Items

Positions

→ Tables



→ Networks

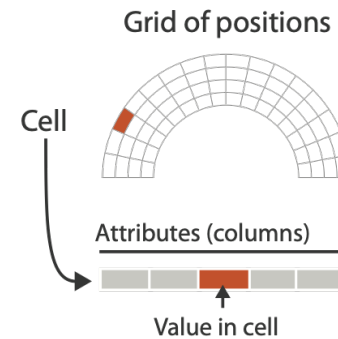


→ Trees



→ Spatial

→ Fields (Continuous)



→ Geometry (Spatial)



Data Semantics

- Start with Data Semantics – what is the meaning of each of the data elements?
- For nominal and ordinal elements what are the unique levels (values)?
- For discrete and continuous elements what are the units and ranges?

Data Wrangling

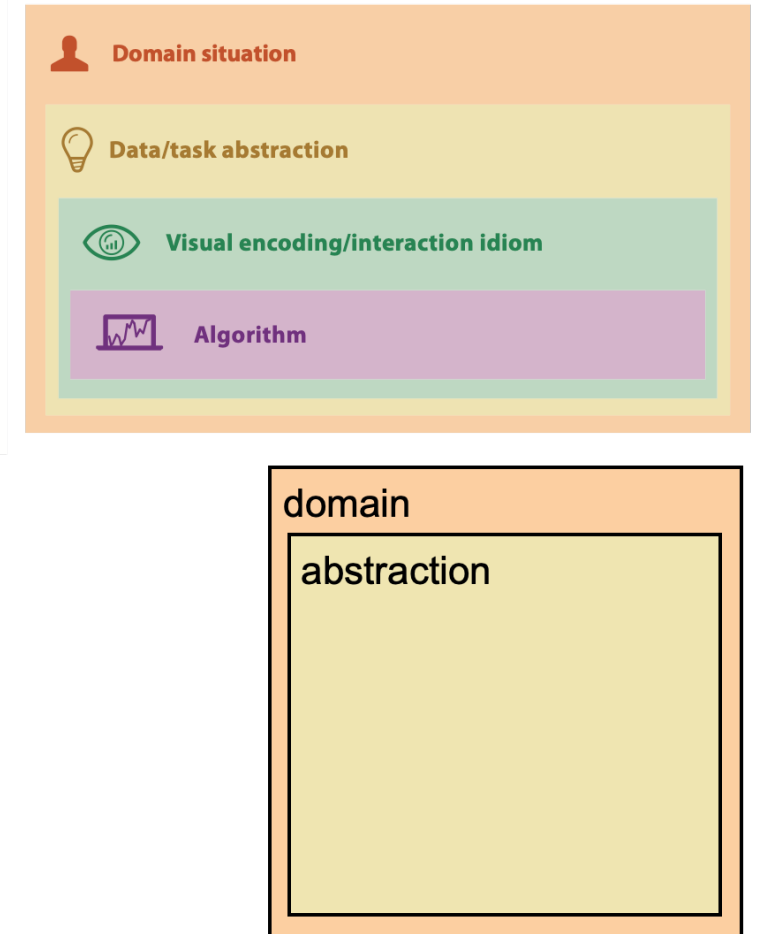
- Missing Data
 - Quantitative Data substitute mean or median
 - Qualitative Data substitute mode
 - Use domain specific function for substitution
- Data in Wrong Format
 - Change format
- Wrong Data
 - Delete data
- Duplicate Data
 - Remove duplicates

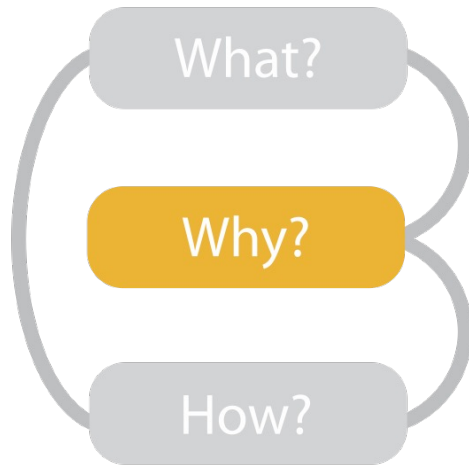
Data Preprocessing

- Data transformation
 - convert from one unit to another (Fahrenheit to Celsius)
- Data binning
 - bin years into decades
- Assigning numerical values to categorical elements
 - grades in class – A:4, B:3, C:2, D:1, F:0

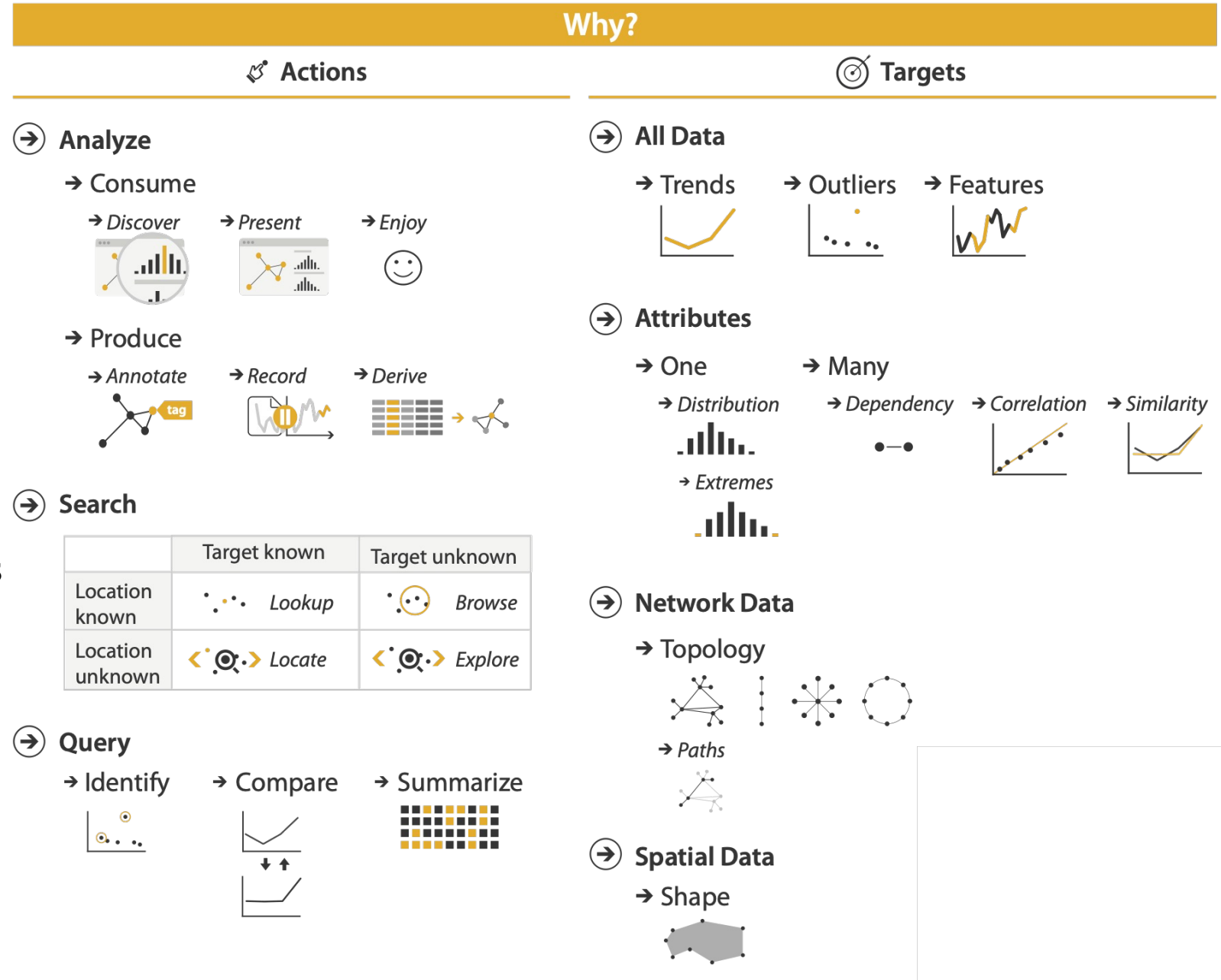
From domain to abstraction

- domain characterization:
details of application domain
 - group of users, target domain, their questions & data
 - varies wildly by domain
 - must be specific enough to get traction
 - domain questions/problems
 - break down into simpler abstract tasks
- abstraction: data & task
 - map *what* and *why* into generalized terms
 - identify tasks that users wish to perform, or already do
 - find data types that will support those tasks
 - possibly transform /derive if need be



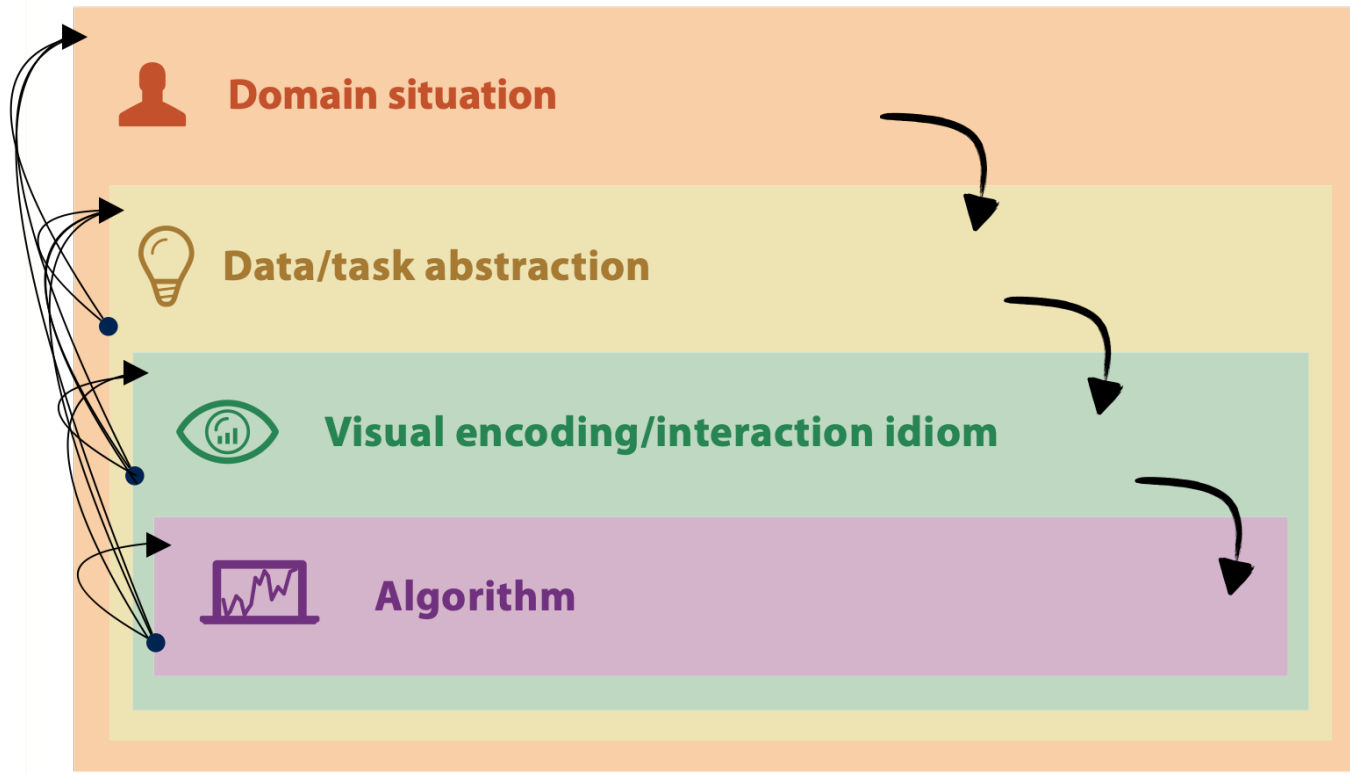


- {action, target} pairs
 - discover distribution
 - compare trends
 - locate outliers
 - browse topology



Nested model

- downstream: cascading effects
- upstream: iterative refinement

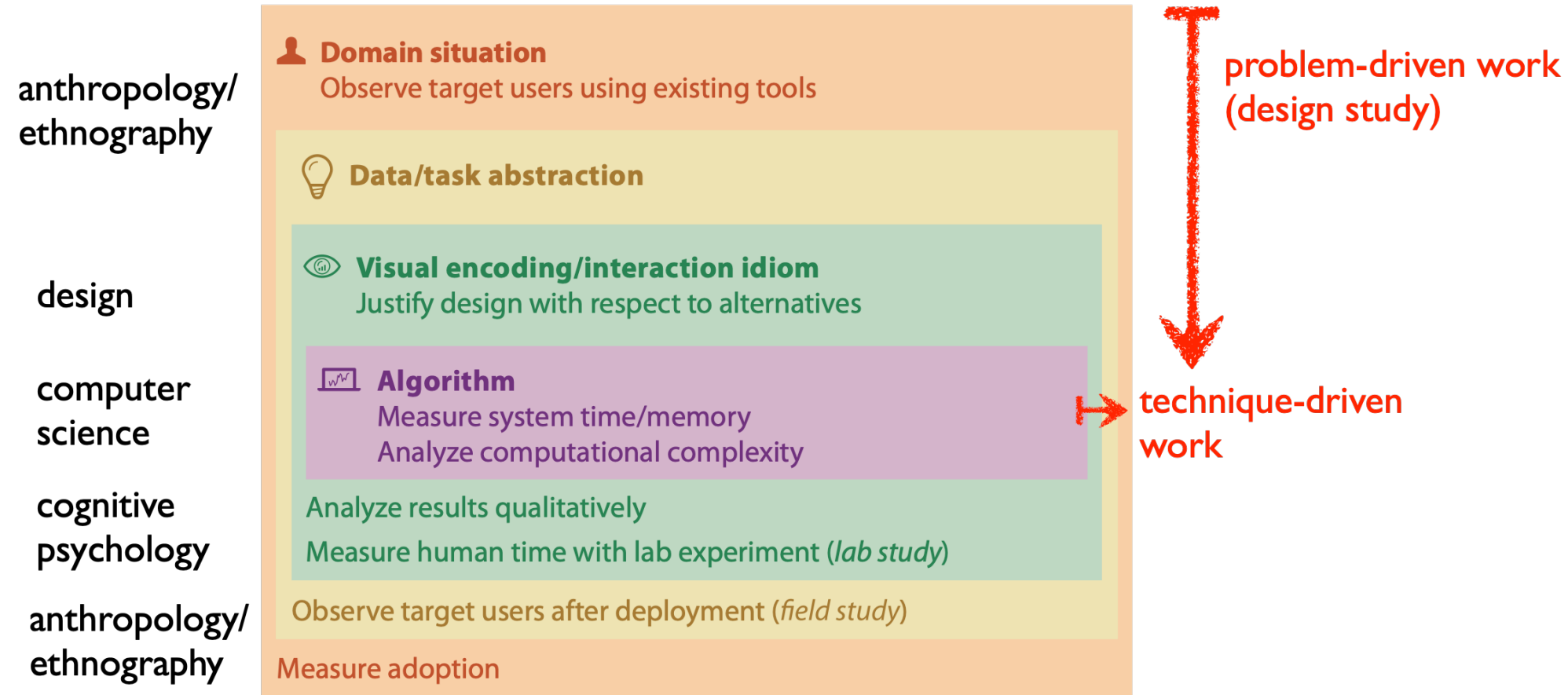


Validation Process

- Start thinking about validation from the very beginning of the design process rather than at the end
- Design process is iterative – a better understanding at one level will refine blocks at other levels
- Domain is used to mean a particular field of interest – microbiology, astrophysics, economics
- Target users are identified and the designer understands the needs of the user

Why is validation difficult?

- solution: use methods from different fields at each level

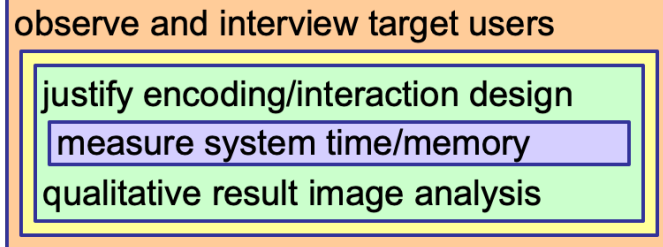


Threats to Validity

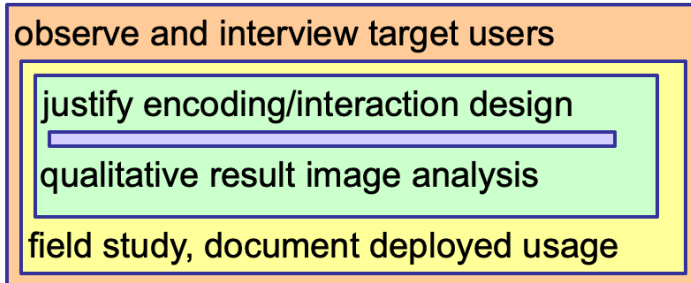
- Domain situation: Wrong problem – you misunderstood their needs
 - field study – observe and interview users in situ
- Data/task abstraction: Wrong abstraction – you are showing them the wrong thing
 - assess how the user deploys the viz tool
- Visual encoding/interaction idiom: Wrong idiom – the way you show does not work
 - justify the design using known perceptual and cognitive principles
- Algorithm: Wrong algorithm – your code is too slow
 - analyze the computational complexity of the algorithm

Analysis examples: Single paper includes only subset of methods

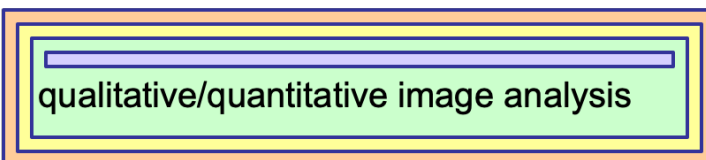
MatrixExplorer. Henry and Fekete. InfoVis 2006.



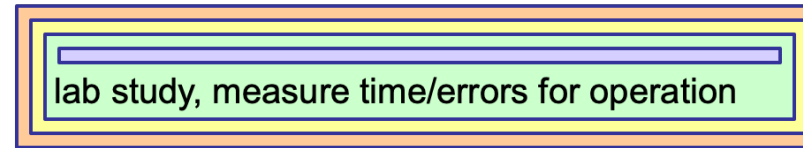
LiveRAC. McLachlan, Munzner, Koutsosios, and North. CHI 2008.



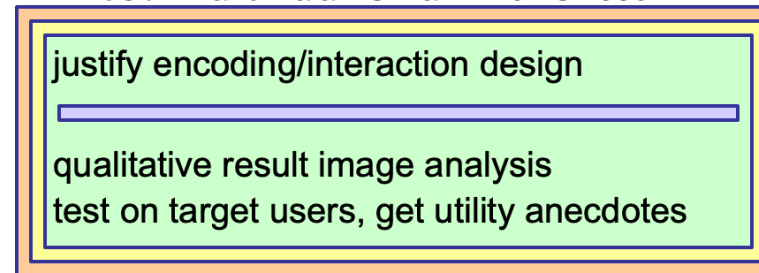
An energy model for visual graph clustering. (LinLog) Noack. Graph Drawing 2003



Effectiveness of animation in trend visualization. Robertson et al. InfoVis 2008.



Interactive visualization of genealogical graphs. McGuffin and Balakrishnan. InfoVis 2005.



Flow map layout. Phan et al. InfoVis 2005.

