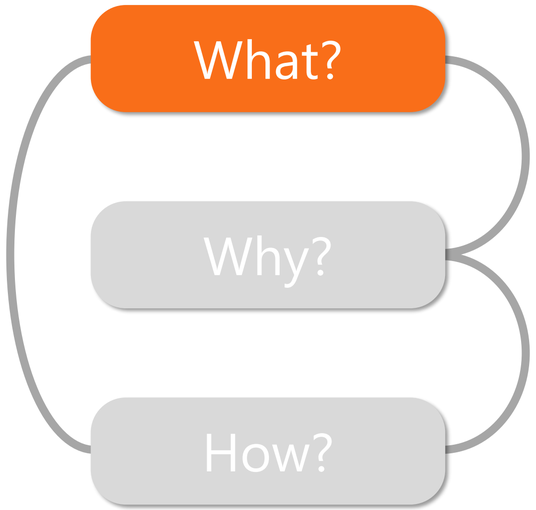
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Visualization of Biological Data – Winter Term 2018/2019

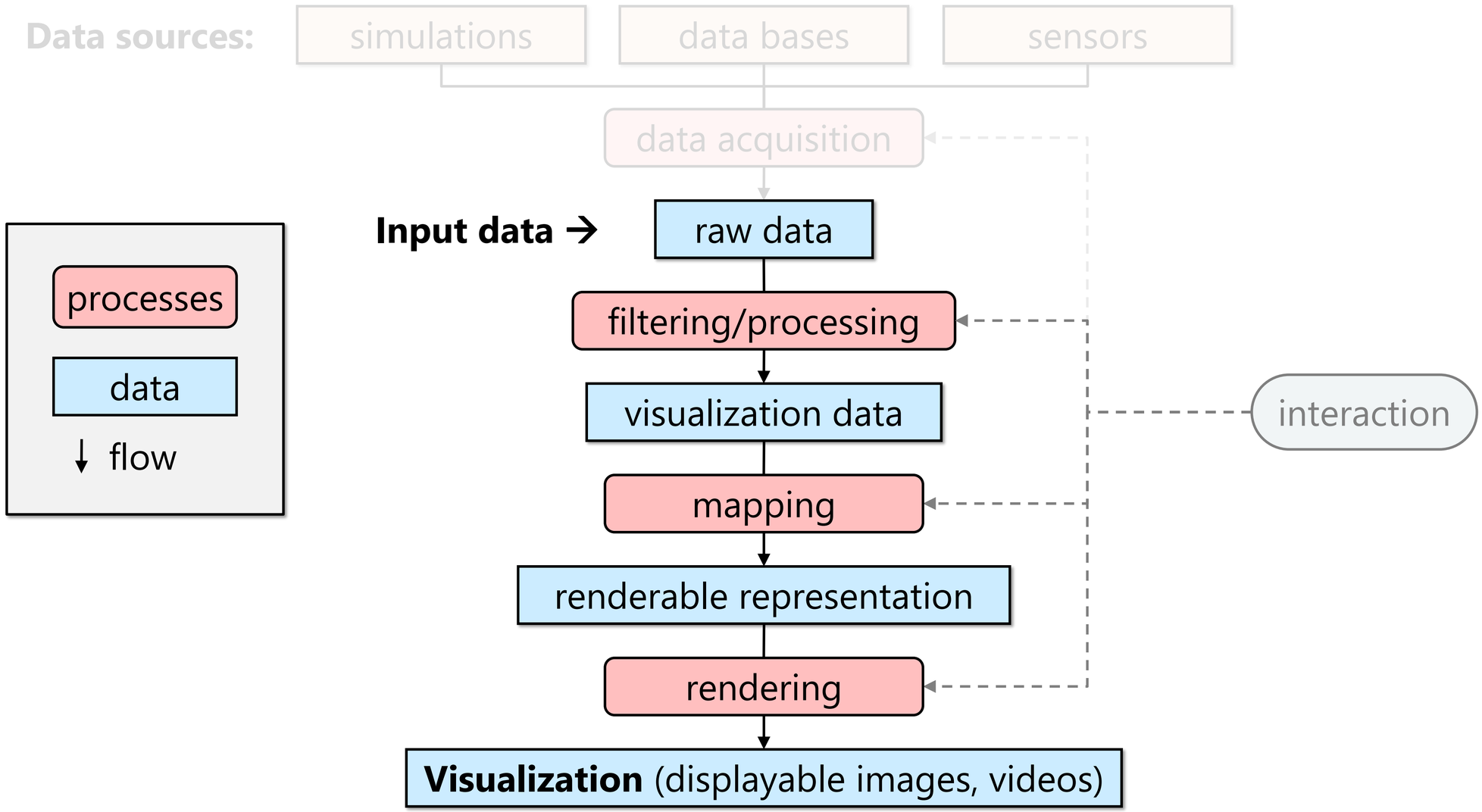
What: Data

Jun.-Prof. Dr. Michael Krone

22.10.2018



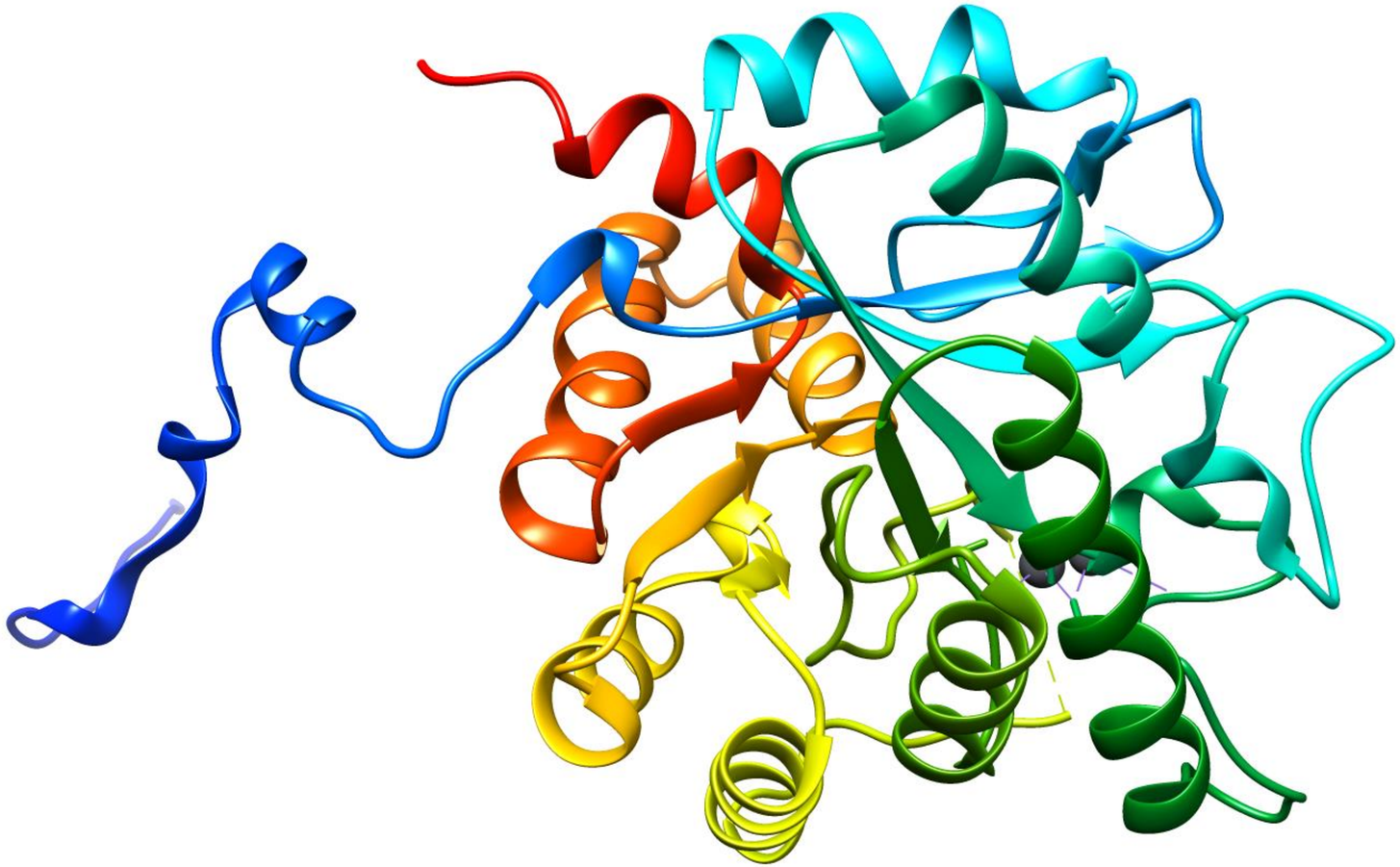
Visualization Pipeline



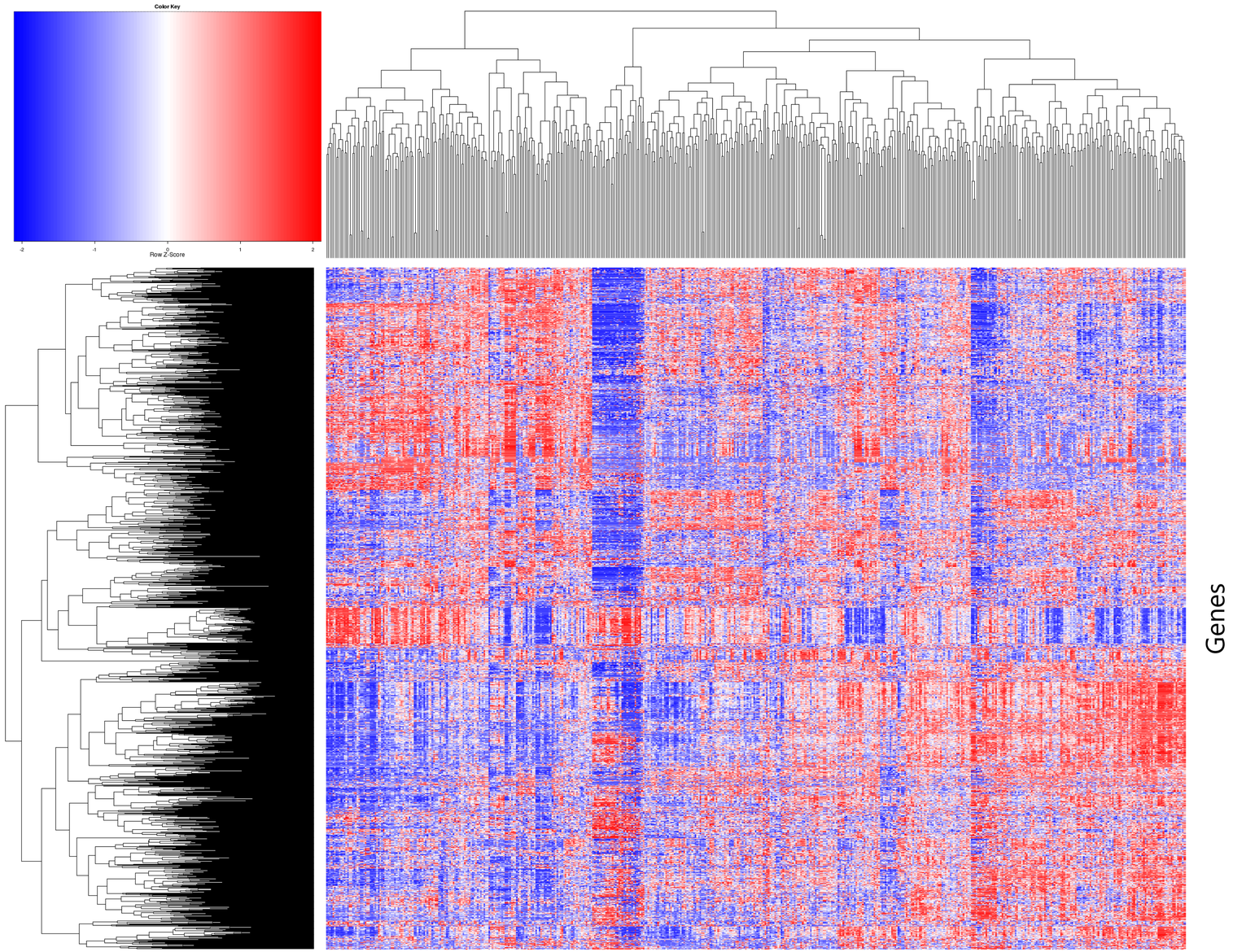
[Haber, McNabb 1990]

What is the input data?

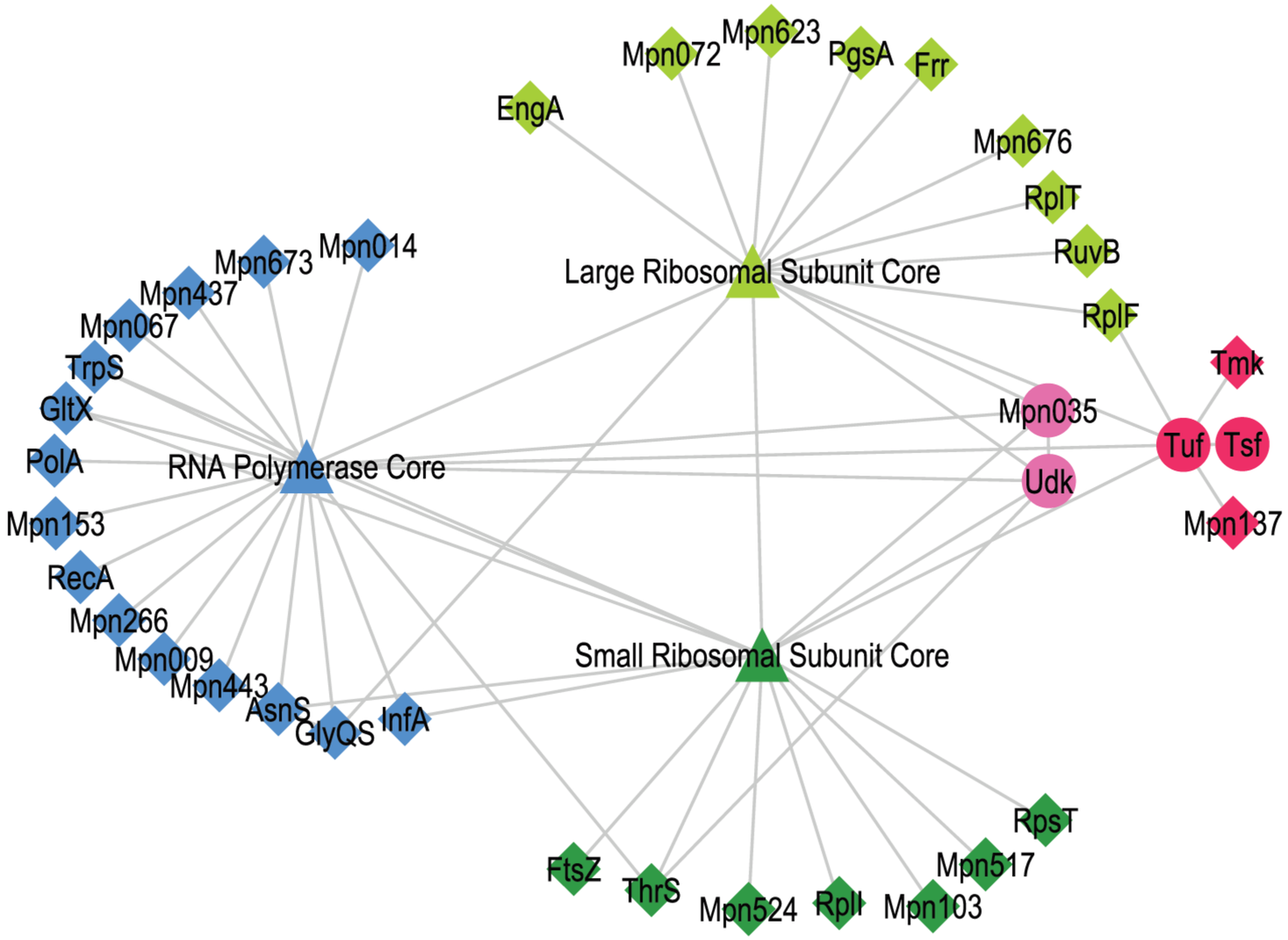
- Example: Visualization of Protein Structure



What is the input data?

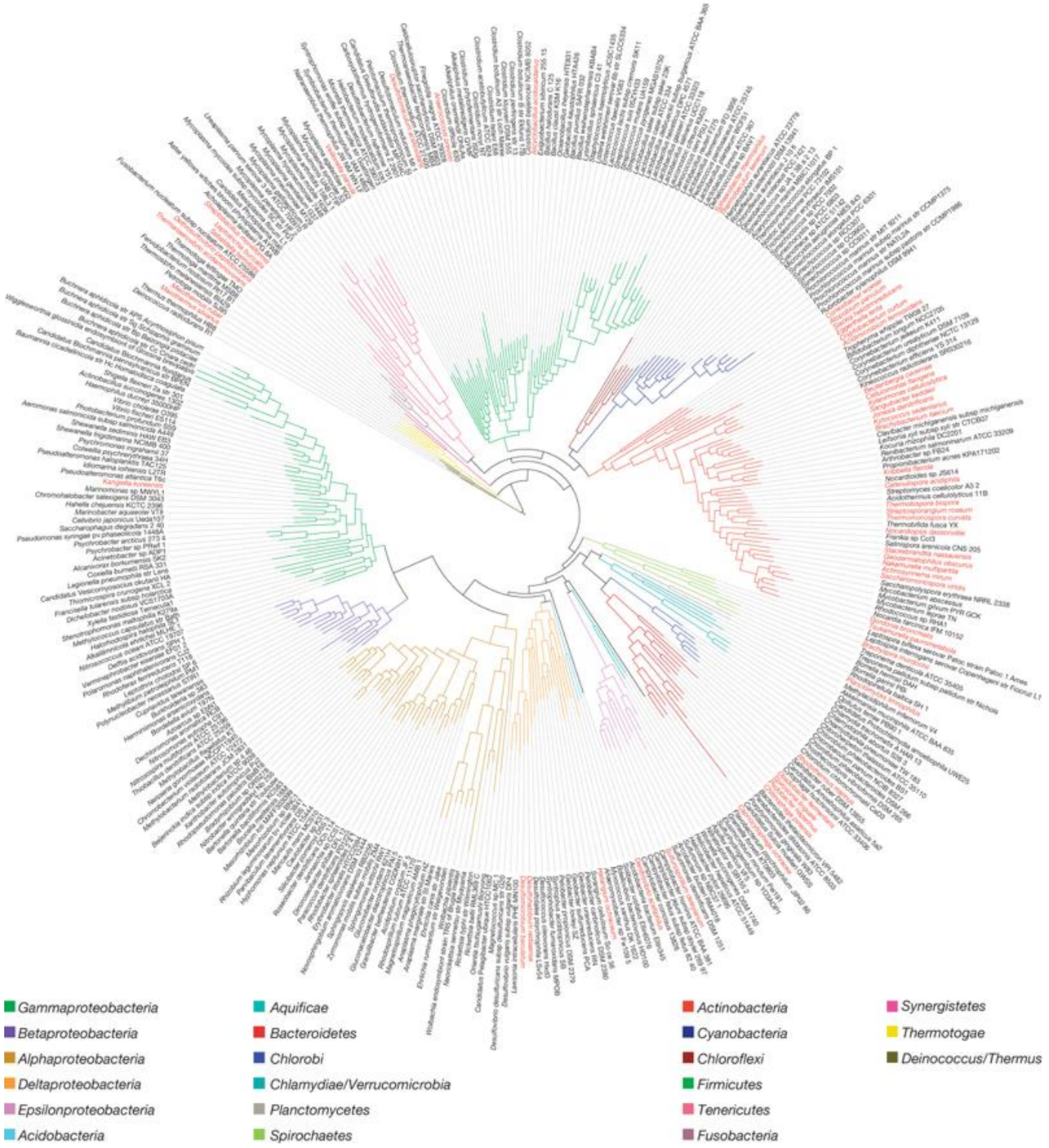


What is the input data?



Gehlenborg et al. 2010, Nature Methods

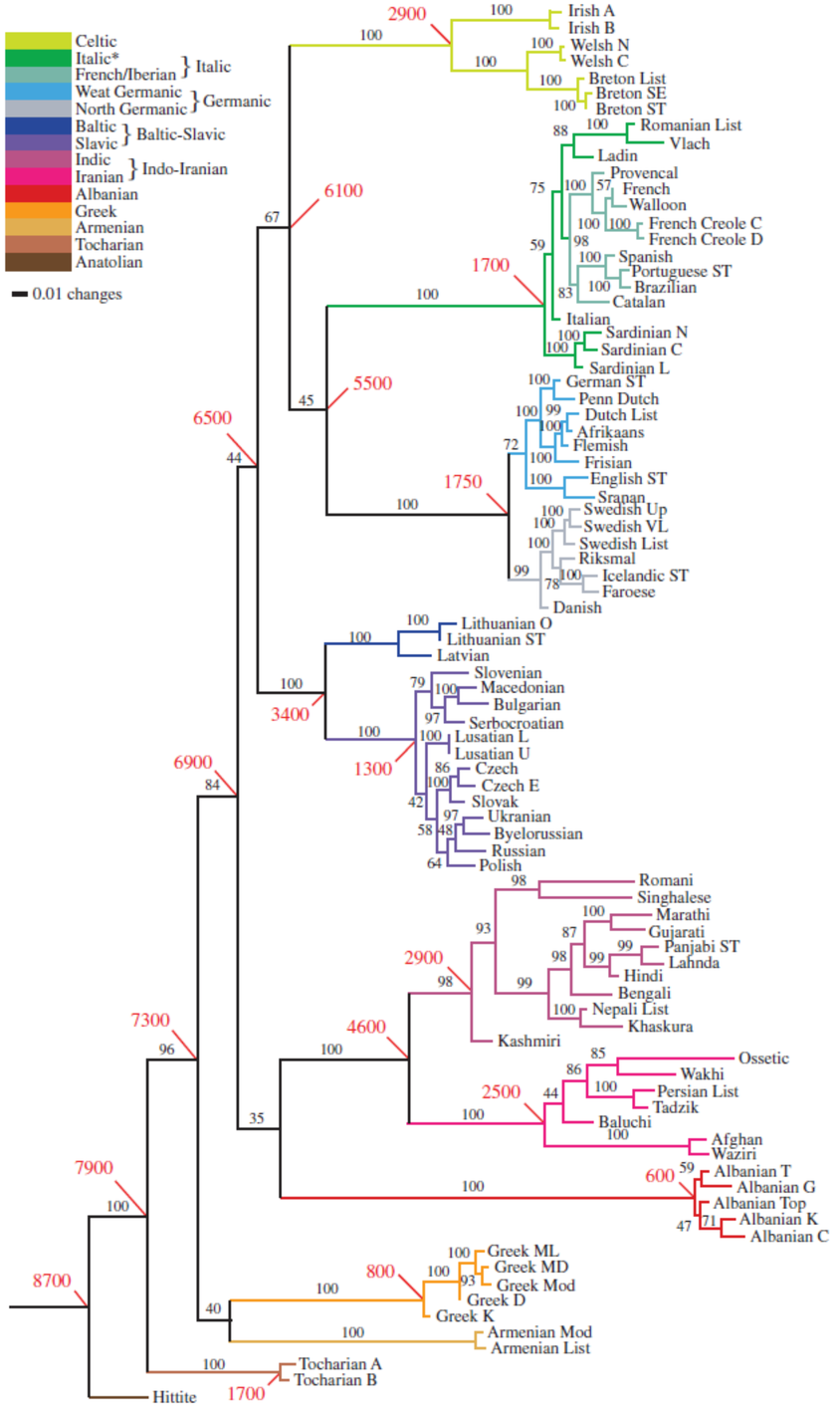
What is the input data?



Wu, D., et al. Nature, 2009

Visualization of

What is the input data?



Heggarty, P., et. al. Philos Trans R Soc London B Biol Sci, 2010

Types

1. define **valid** operations on items

5.3	+	3.1	=	8.4	Quantitative
Alice	+	Bob	=	?	Categorical

Types

1. define **valid** operations on items
2. define **structure** of items

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Munzner, 2014

Semantics

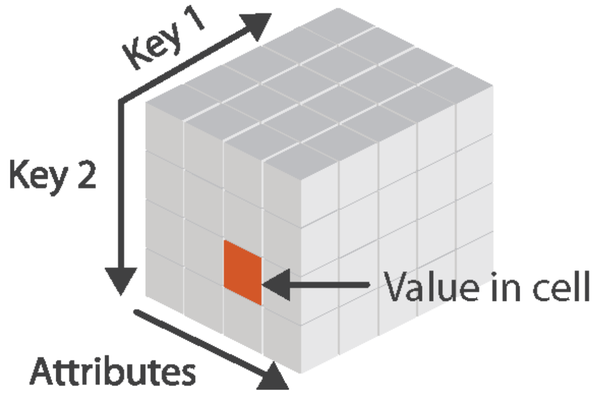
1. define **meaningful** operations on items

5.3	+	3.1	=	8.4	Quantitative
Alice	+	Bob	=	?	Categorical
72076	+	88048	=	160124?	Categorical!

Semantics

1. define **meaningful** operations on items
2. distinguish **keys** from **values**

➔ *Multidimensional Table*



Munzner, 2014

What?

Datasets

Data Types

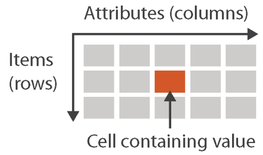
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

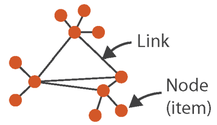
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

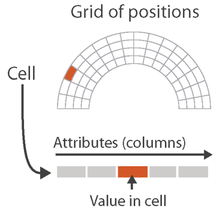
- Tables



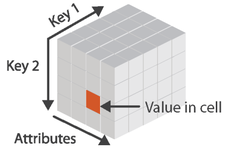
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

Dataset Availability

- Static

- Dynamic

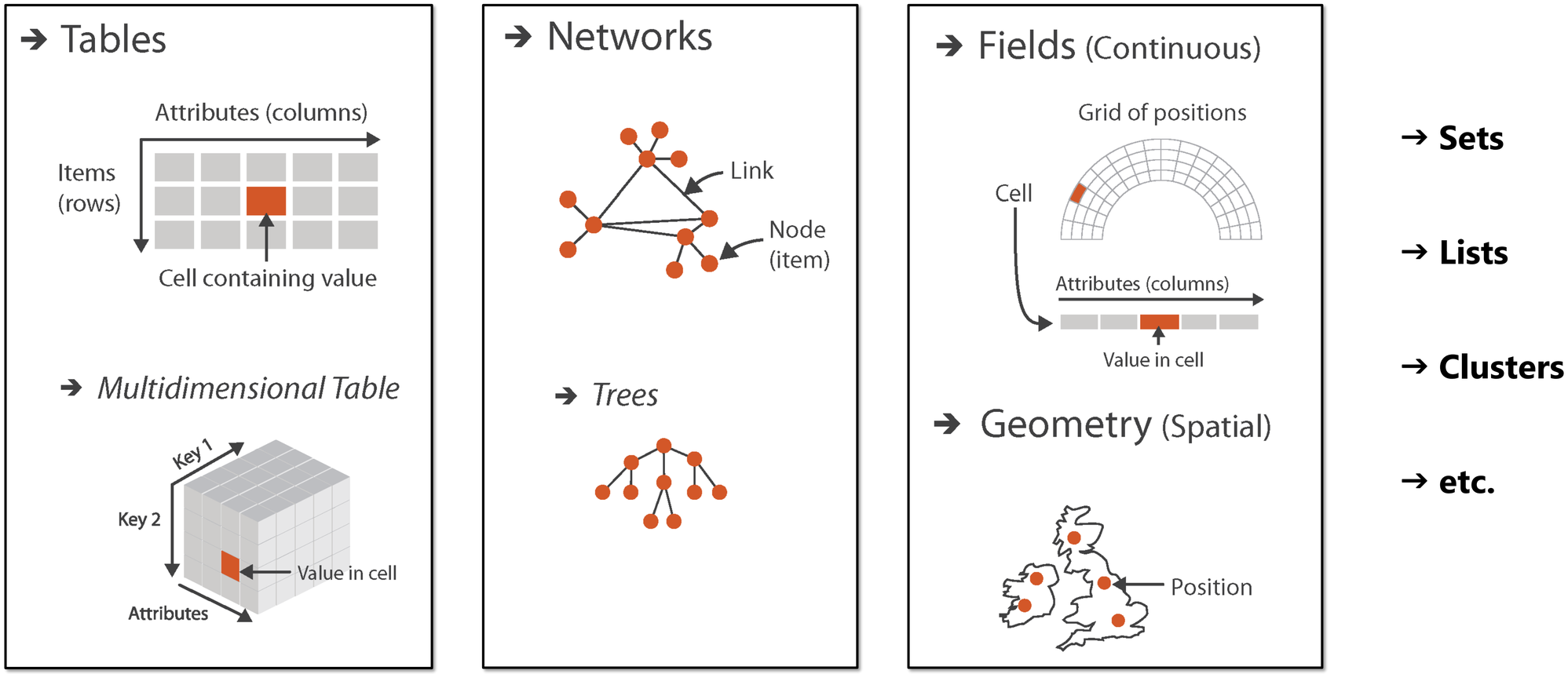
Data Types

Data Types

- Items (entities)
- Attributes (properties)
- Links (relationships)
- Grids (continuous fields)
- Positions (locations)

Dataset Types

Dataset Types



➔ **Sets**

➔ **Lists**

➔ **Clusters**

➔ **etc.**

Munzner, 2014

- Visualization vs. computer graphics
  - Geometry (i.e., geometric representation in visualization) is design decision!

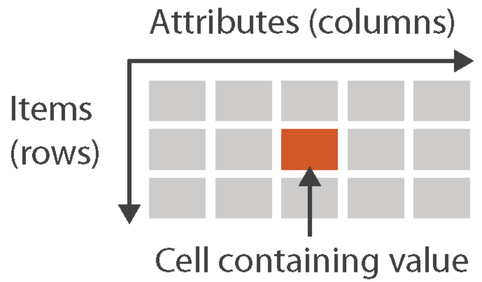
Tables

Cars marketplace				
vendor	Model	Price	Mileage	VIN Code
Chevrolet	Corvette	17226	25965.0	ILLAKAWAZDZ
Chevrolet	Corvette	34229	46429.0	RCPNSRYGXOI
Chevrolet	Corvette	27982	50209.0	NWLGCEVEHGI
Chevrolet	Corvette	51825	72998.0	NGVZSCIZGSM
Chevrolet	Corvette	52845	34364.0	PSDRUYYOIJG.
Chevrolet	Malibu	37874	37273.0	VLFPQPWNEFD
Chevrolet	Malibu	15600	71441.0	EXLJGDWOZSA
Chevrolet	Malibu	52447	46700.0	NLMGJZAKBRD
Chevrolet	Malibu	27129	36254.0	OIPFUIENLEHSX
Chevrolet	Malibu	28846	77162.0	WRCOOFREZLL
Chevrolet	Malibu	46165	60590.0	HUFTTHQHSFJF
Chevrolet	Malibu	18263	37790.0	JLMHNAFSHVD

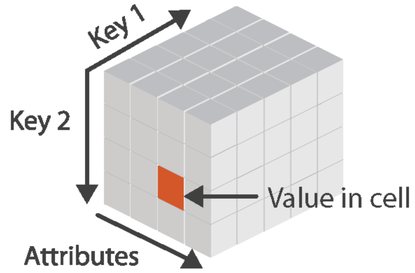
Tables

- Very common dataset type:
  - row is an item
  - column is an attribute
- Can be visualized in various ways

➔ Tables



➔ *Multidimensional Table*

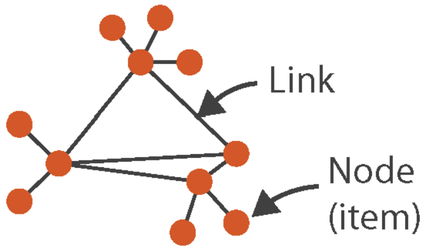


Munzner, 2014

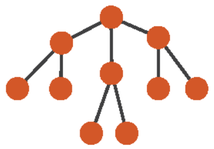
Networks

- To specify relationships between items (nodes)
- In a network (as a dataset) nodes do not have specific positions
- Trees have hierarchical structure and no cycles

➔ Networks



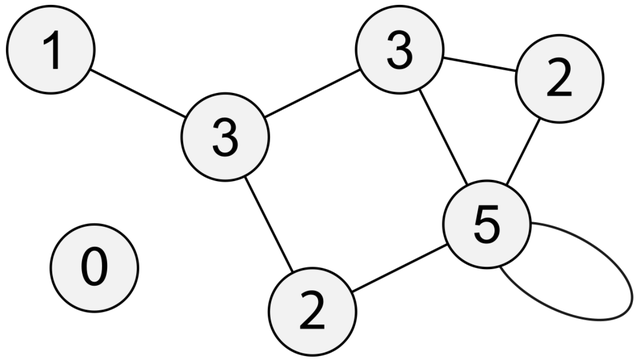
➔ *Trees*



Munzner, 2014

Graphs

- 3.1 Definition
  - A graph is an ordered pair G = (V, E) consisting of
    - a set V of vertices and
    - a set E of edges that is a subset of V x V
  - To avoid ambiguity, this type of graph may be described precisely as undirected and simple.



wikipedia.org

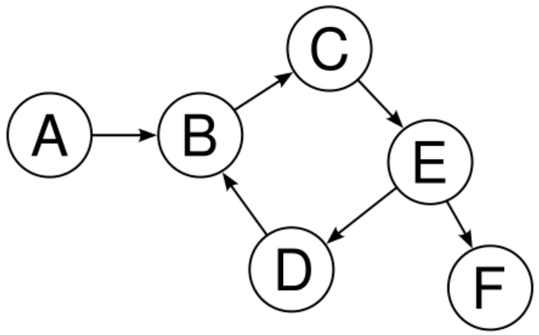
Graphs

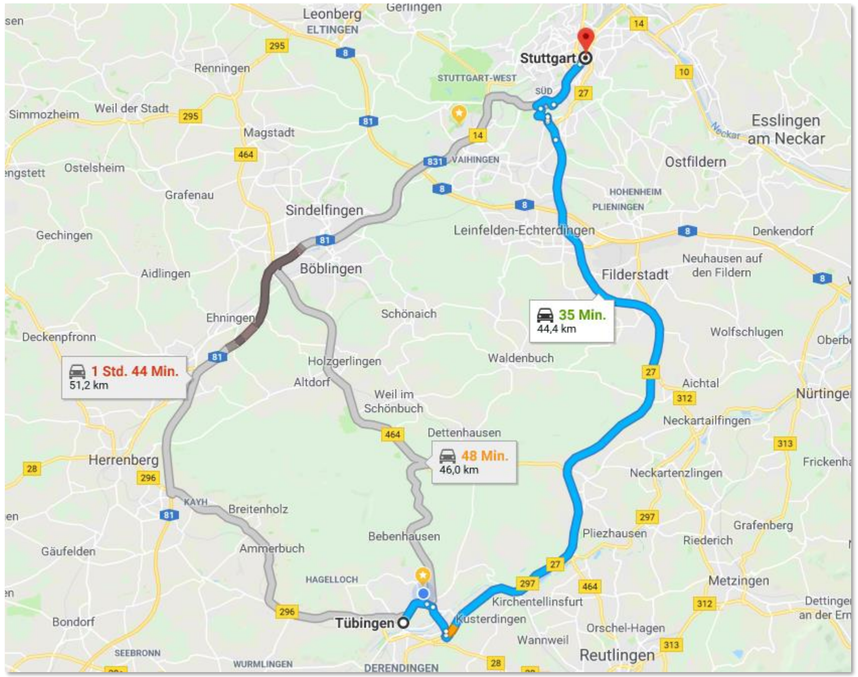
- 3.1 Definition and Data Model
  - Leonard Euler:
    - Seven Bridges of Königsberg (1735):

→ Foundation of Graph Theory

Networks

- 3.1 Definition and Data Model
  - Network:
    - A directed graph (digraph) with weighted edges in the context of graph theory is called a network. Network analysis has many practical applications, for example, to model and analyze traffic networks.

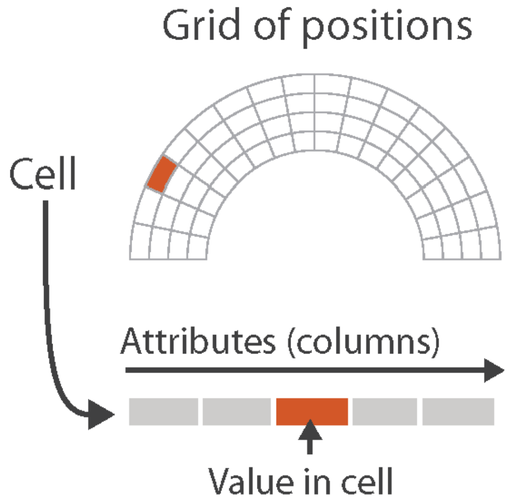




Fields

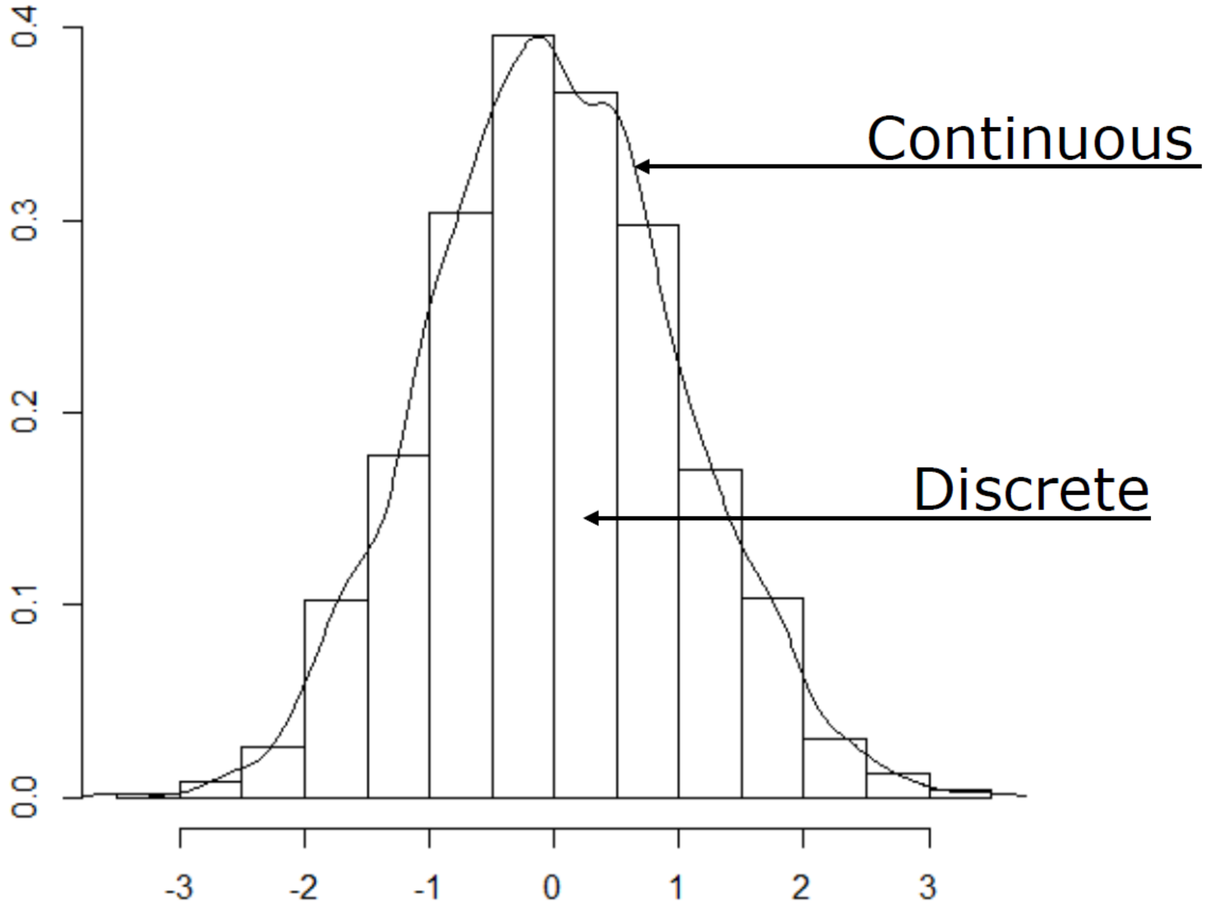
- Each **cell** contains measurements or calculations from a **continuous domain**.
  - e.g., temperature, density, pressure, speed, etc.
- Sampling frequency and interpolation (estimating values between samples).

➔ Fields (Continuous)

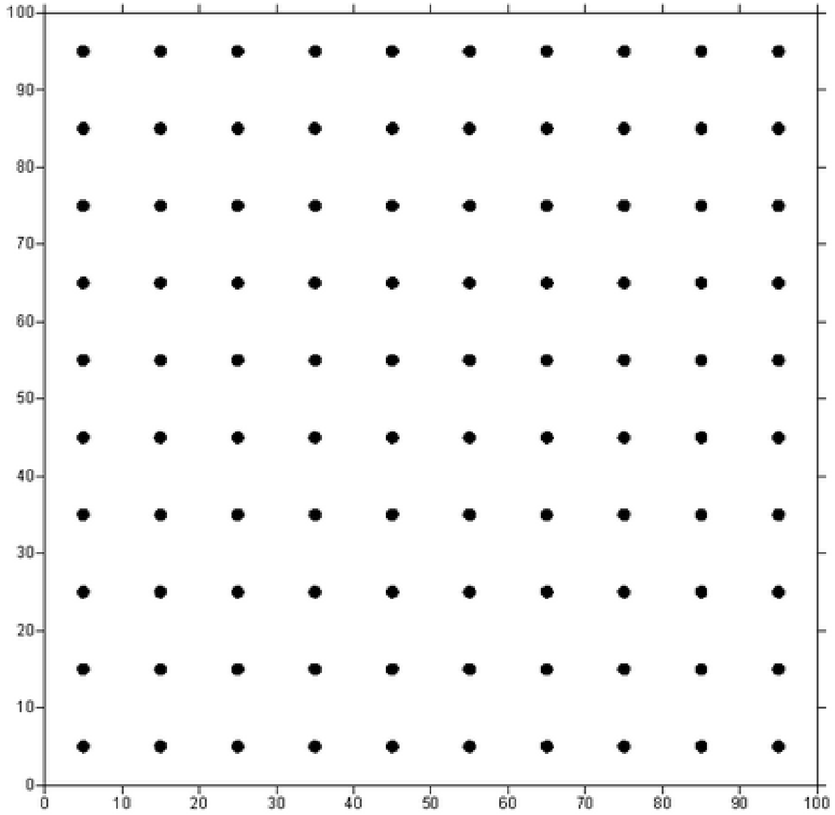


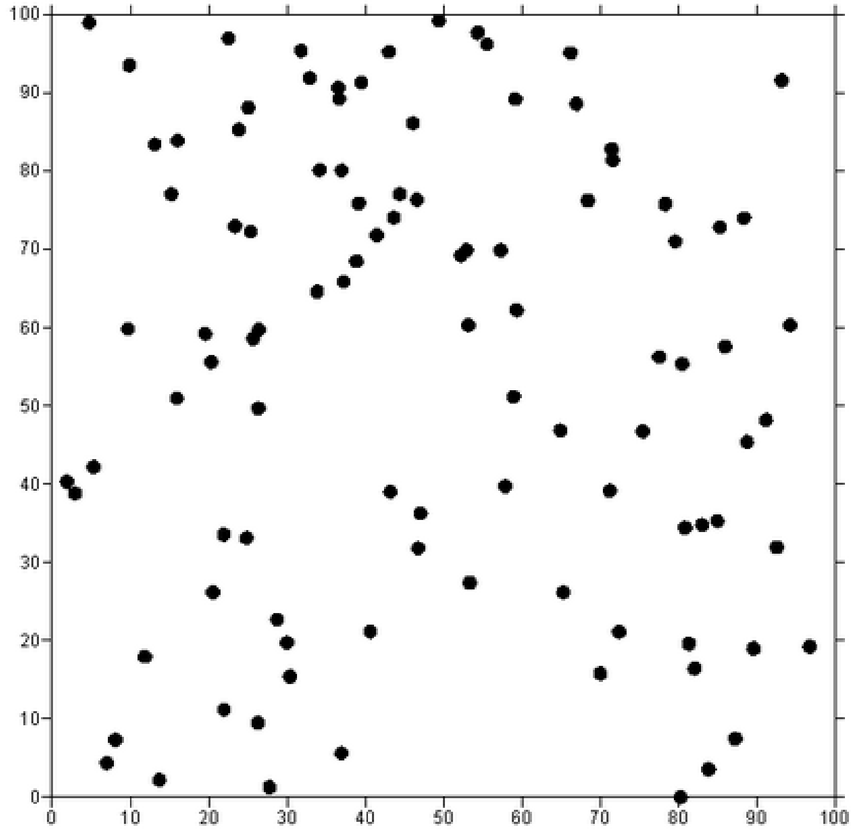
Munzner, 2014

Continuous Domain



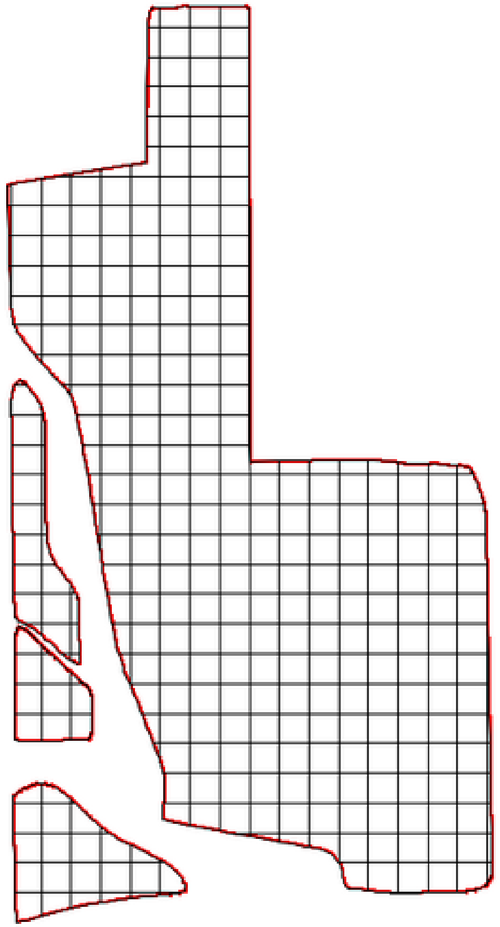
Spatial Fields



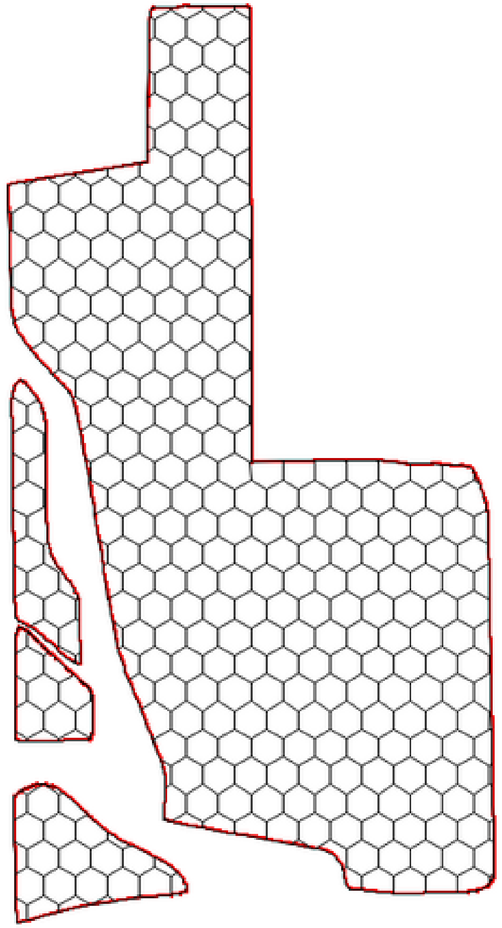


www.spatialanalysisonline.com

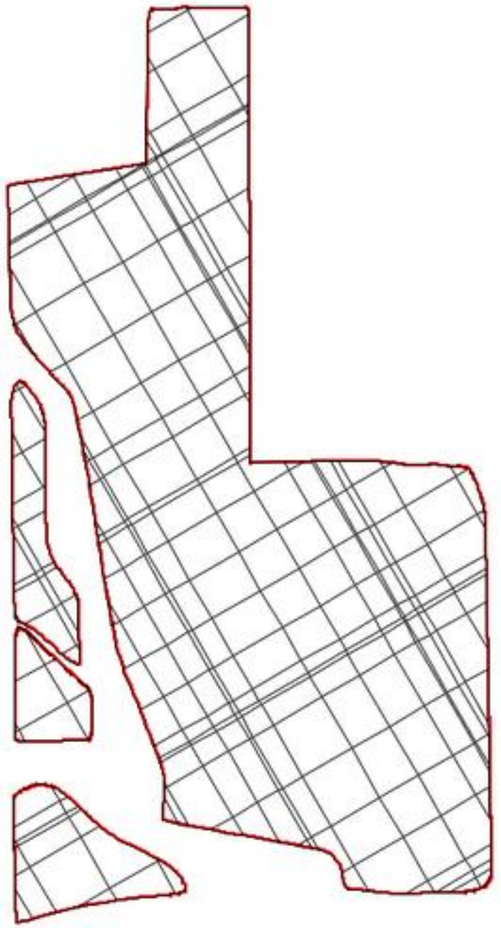
Grid Types



Rectangular



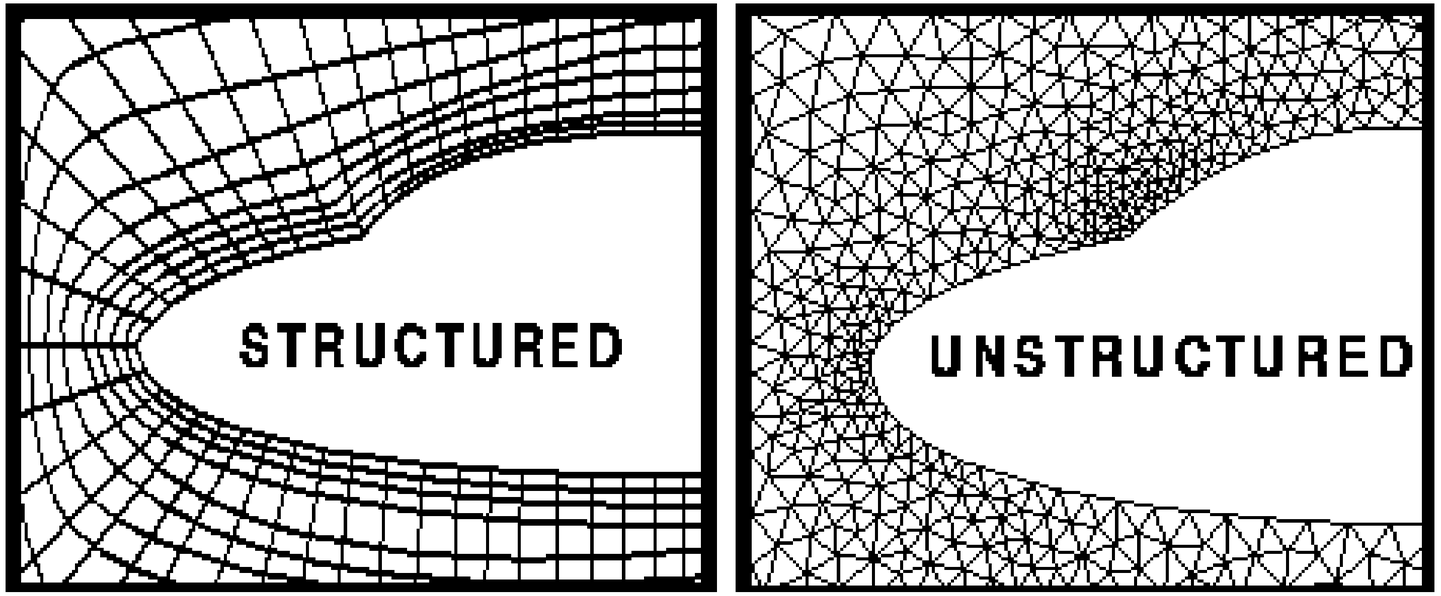
Hexagonal



Random rectangular

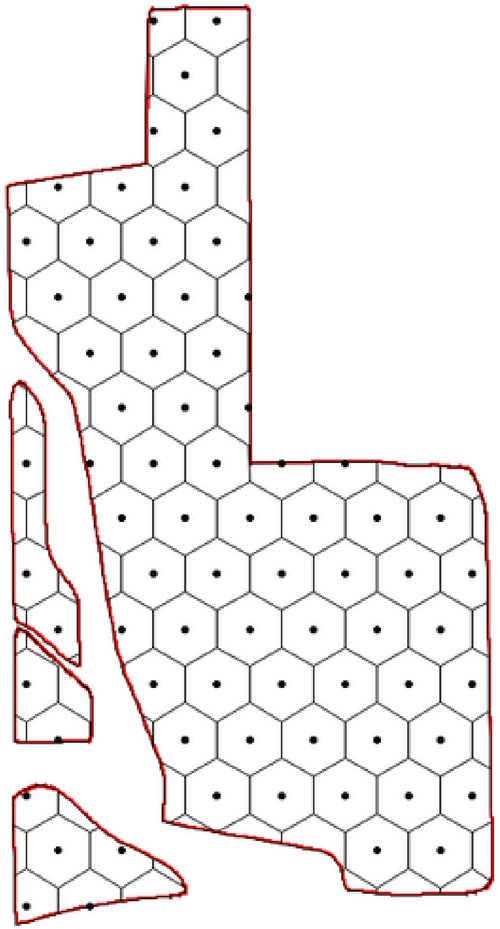
www.spatialanalysisonline.com

Grid Types

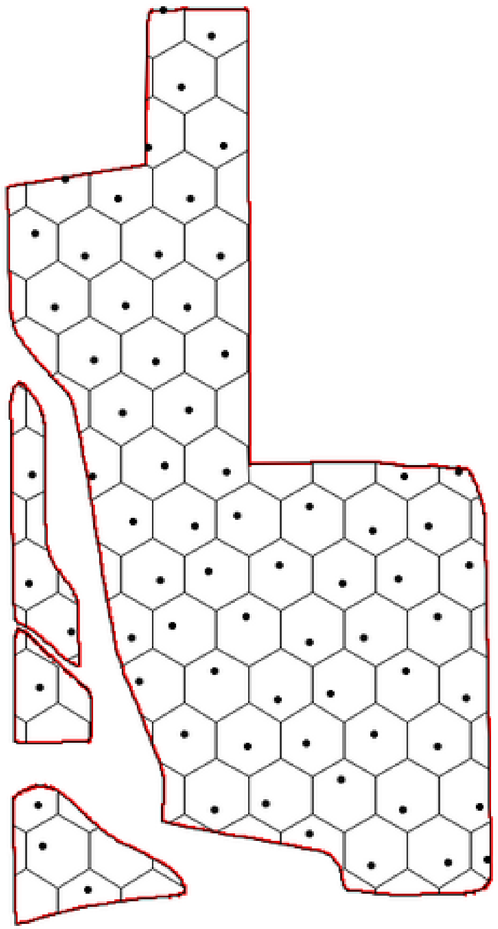


https://www.nas.nasa.gov/Software/FAST/RND-93-010.walatka-clucas/htmldocs/chp_16.surferu.html

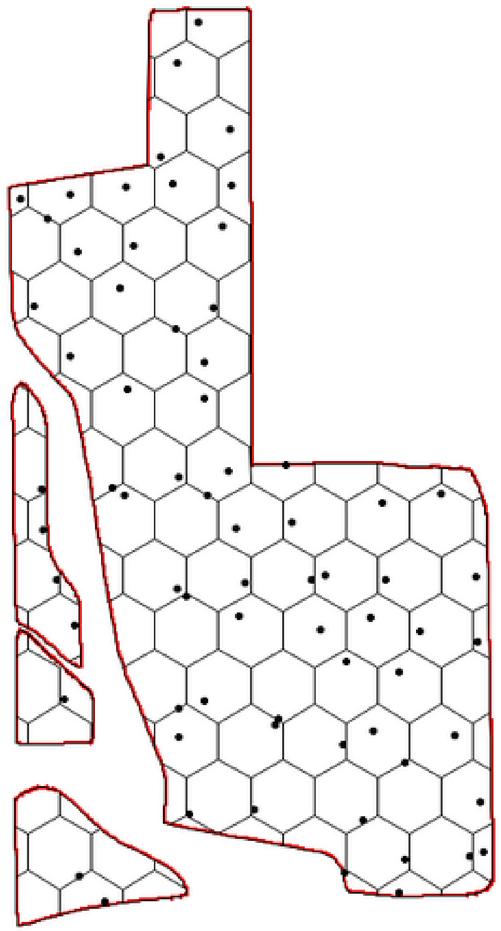
Grid Sampling



Regular



Systematic



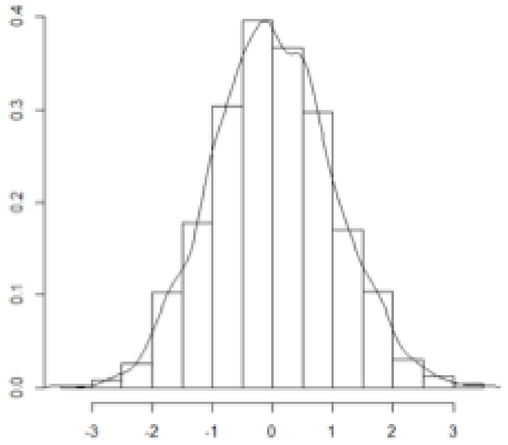
Random

www.spatialanalysisonline.com

Spatial Field Types

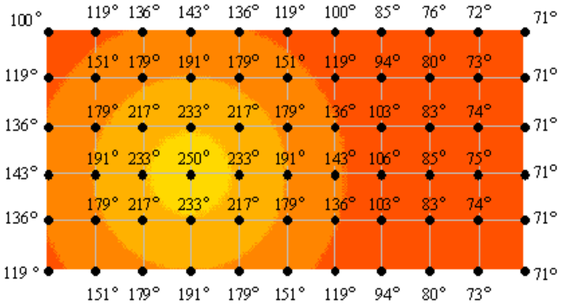
- Scalar (one attribute per cell)
  - e.g., temperature in the room
- Vector (two or more attributes per cell)
  - e.g., velocity of air in the room (direction + speed)
- Tensor (many attributes per cell)
  - e.g., stress (forces in three orthogonal directions)

Scalar Fields



1D Domain

1d1v



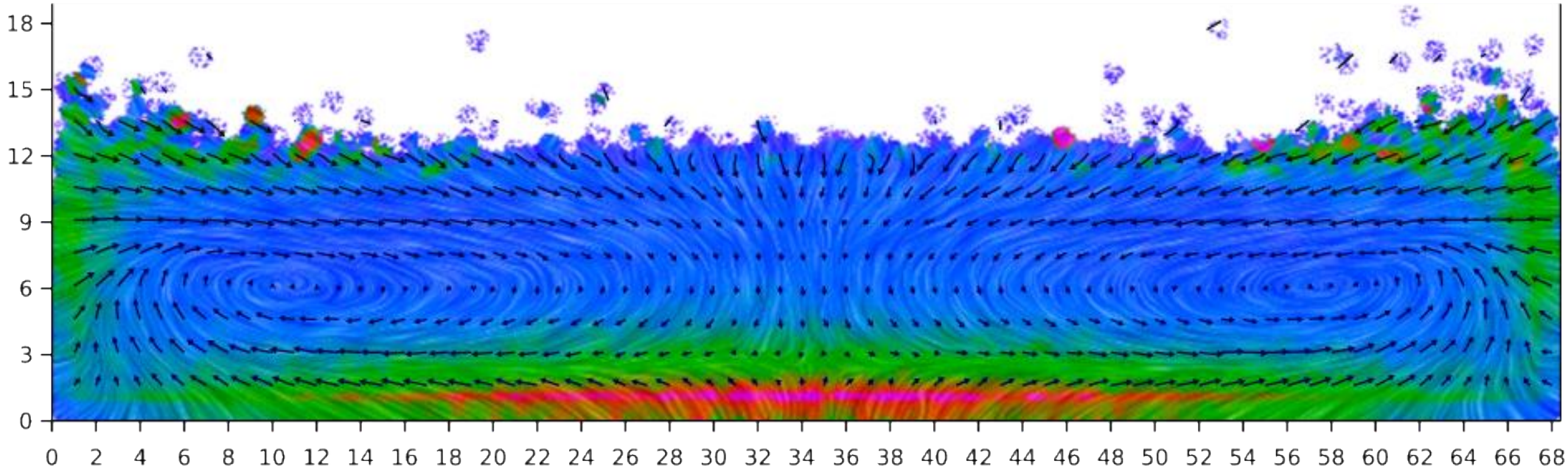
2D Domain

2**d**1**v**

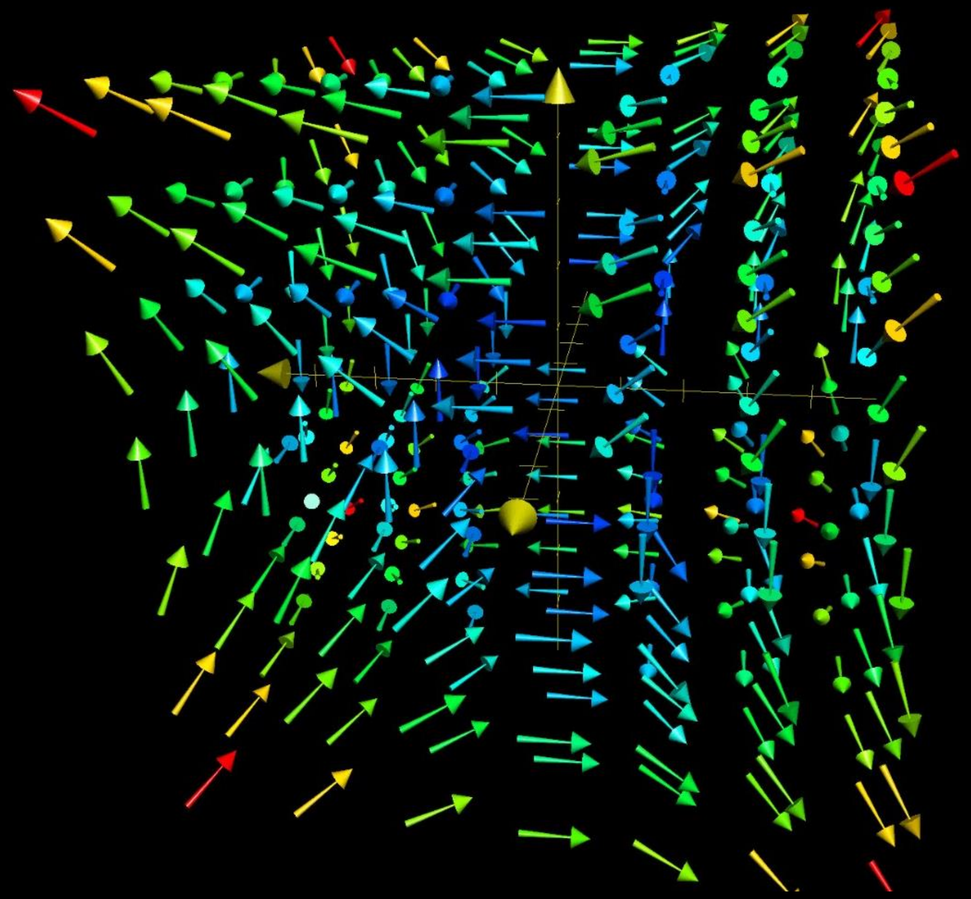
3D Domain

3**d**1**v**

Vector Fields

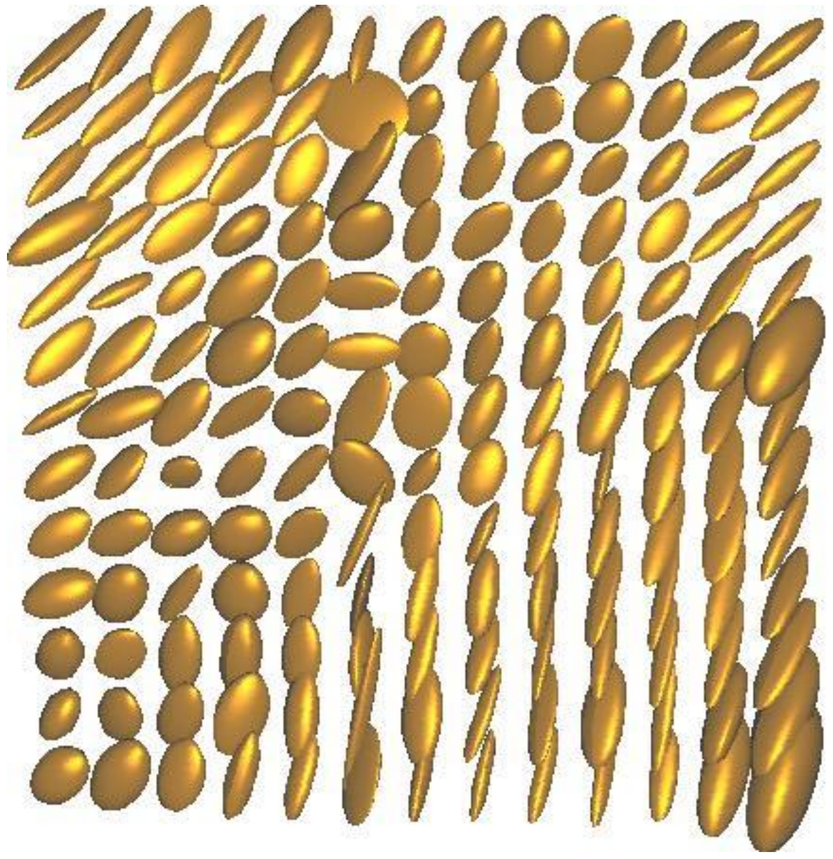


2**d**2**v**

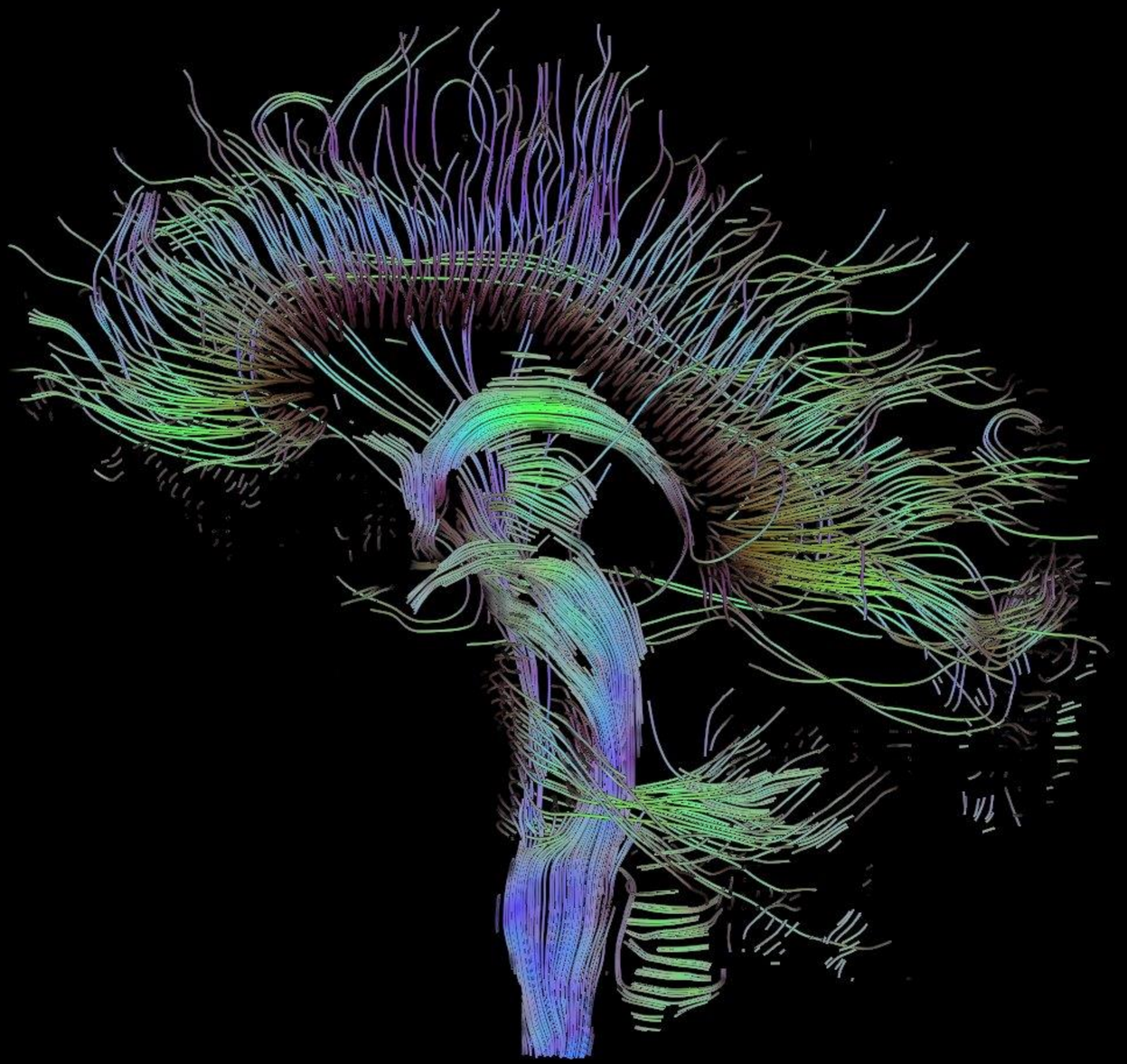


3**d**3**v**

Tensor Field



3**d**9**v**

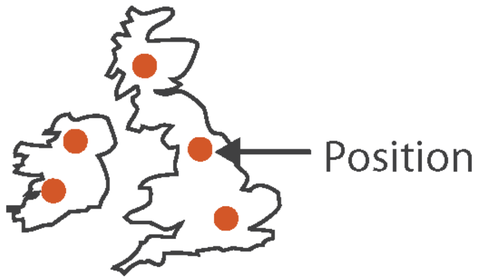


3**d**9**v**

Geometry

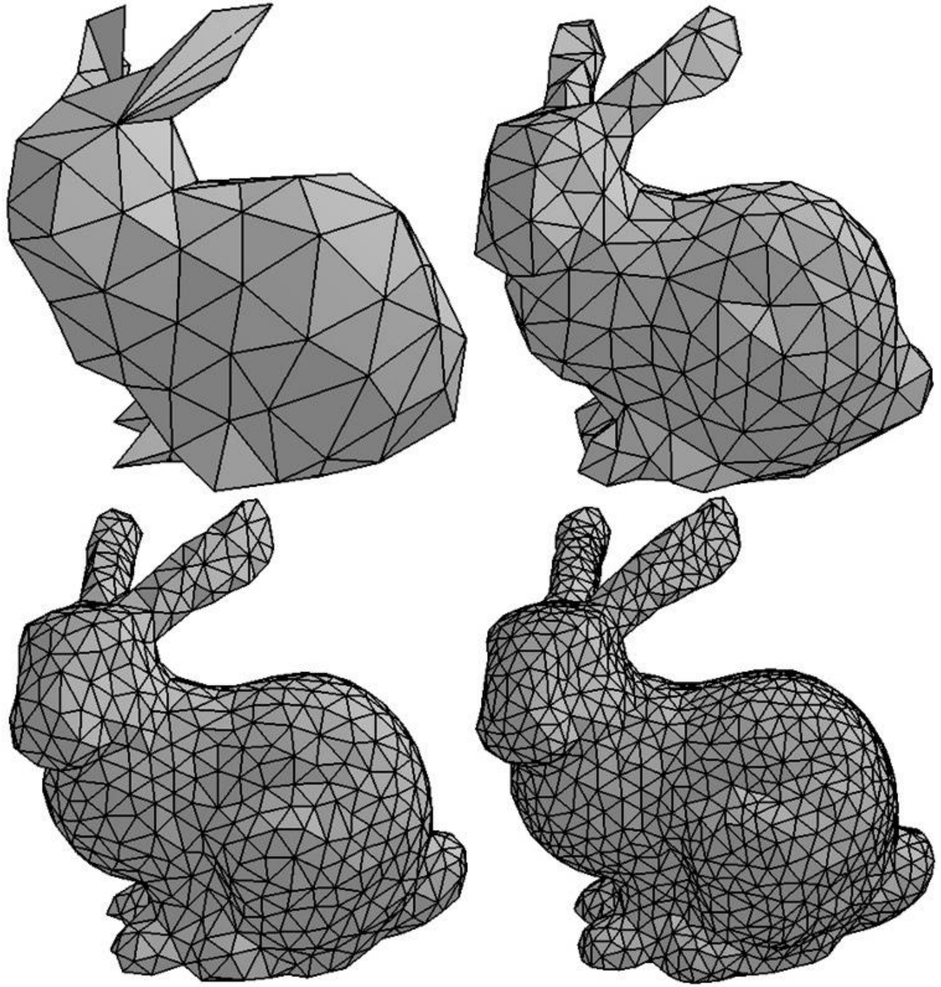
- Item shape + spatial location
  - Points, lines, curves, etc.
- Shape understanding.
- May include hierarchical structure or material properties.
- Not so interesting for data visualization (computer graphics).

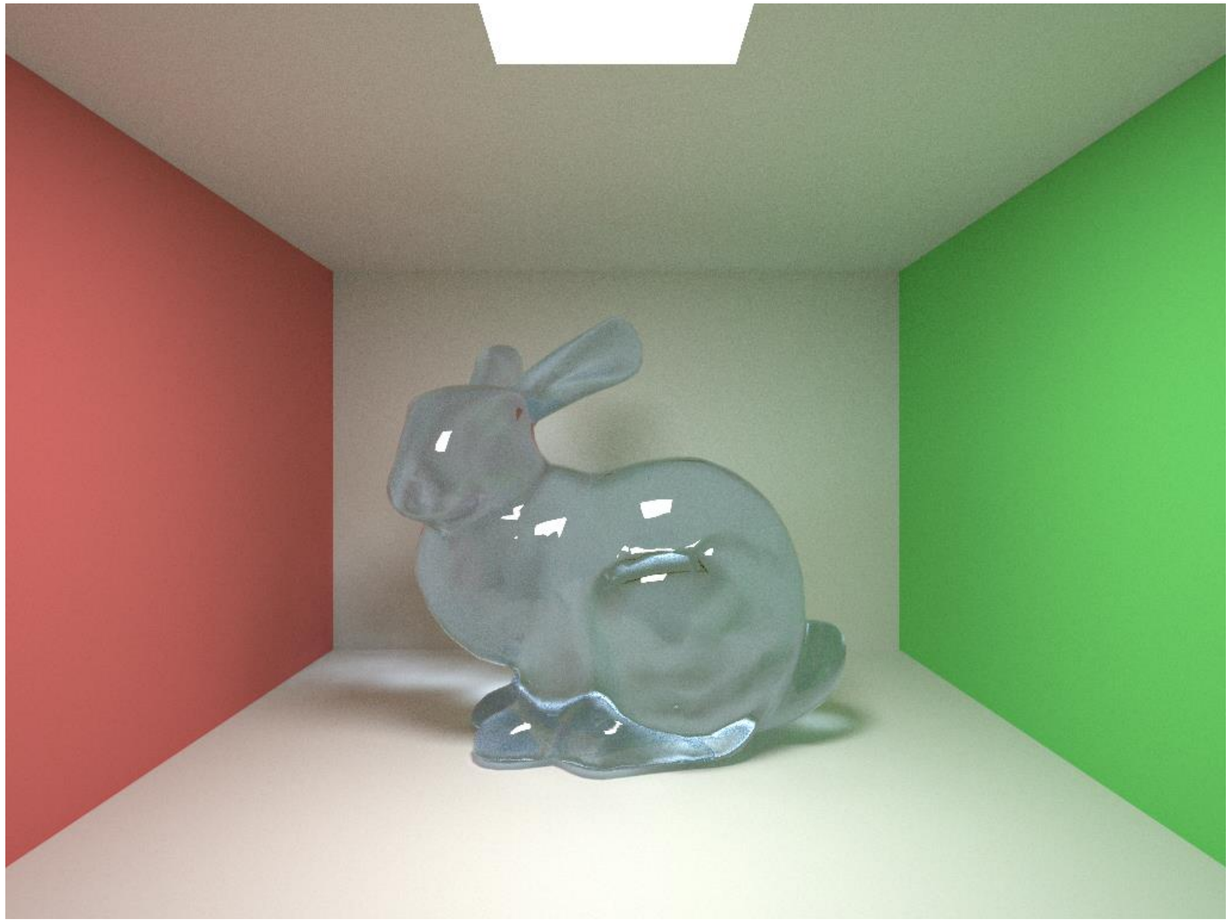
➔ Geometry (Spatial)



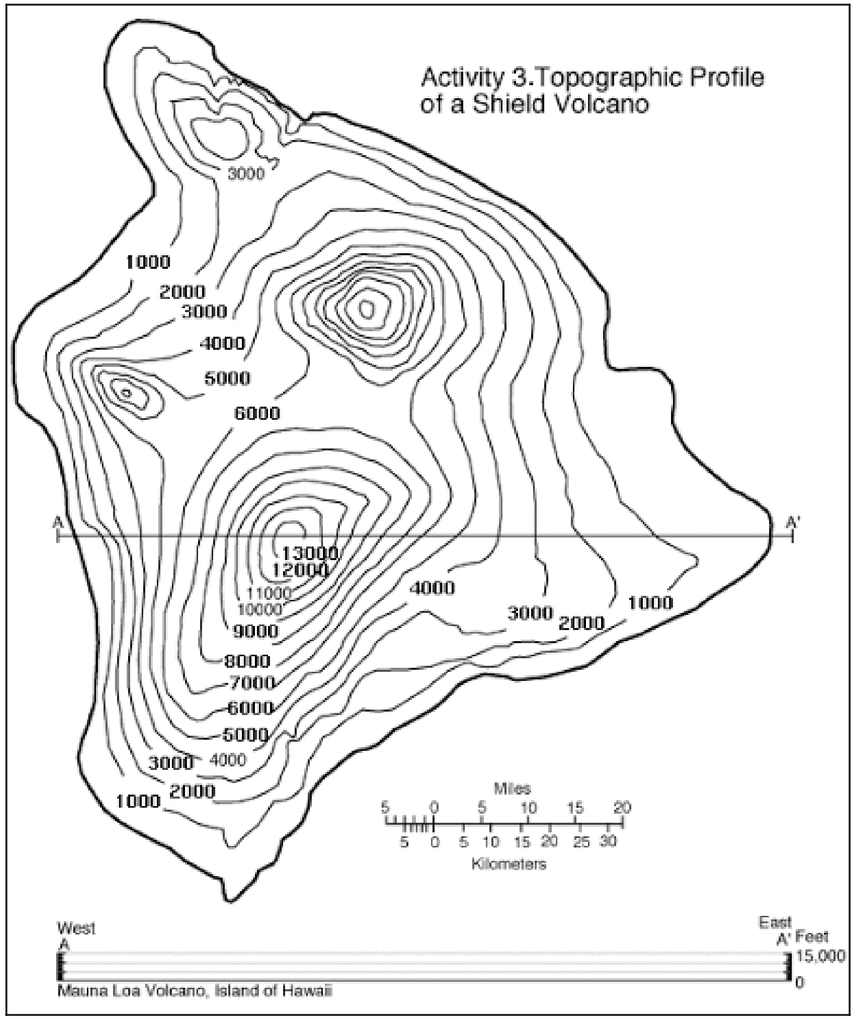
Munzner, 2014

Geometry





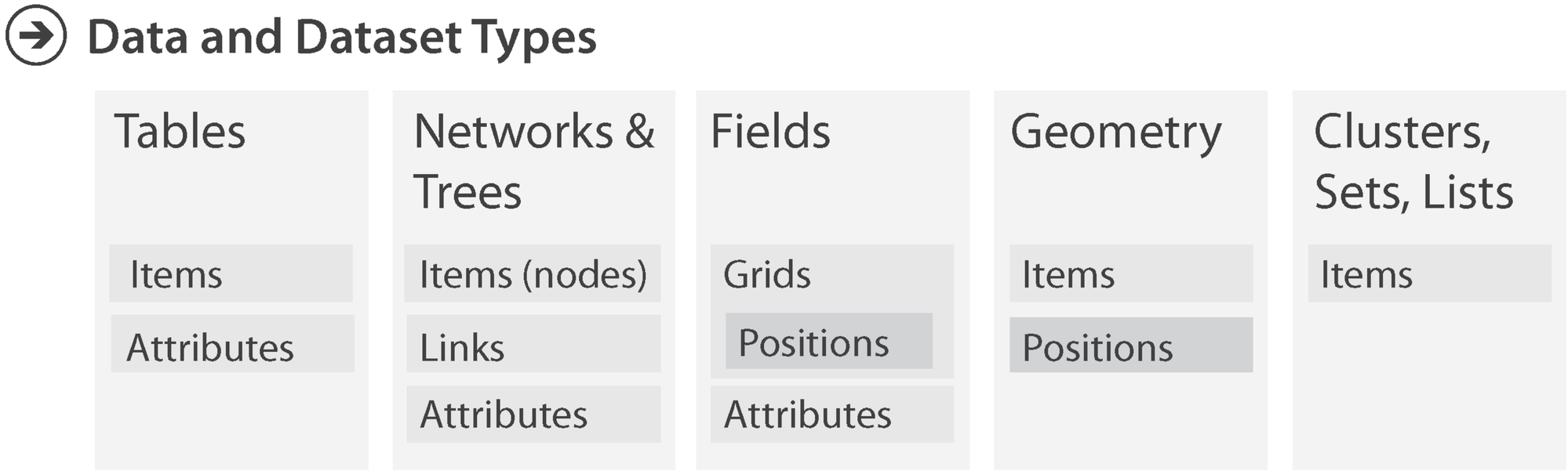
Contours


Activity 3.Topographic Profile
of a Shield Volcano
3000
1000
2000
3000
4000
5000
6000
A
A'
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
4000
3000
2000
1000
Miles
5 0 5 10 15 20
5 0 5 10 15 20 25 30
Kilometers
West
A
East
A' Feet
15,000
0
Mauna Loa Volcano, Island of Hawaii

Other Datasets

- Sets
- Lists
- Clusters
- Paths (in a network)
- Compound networks (network + tree)

Data and Dataset Types



Munzner, 2014

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Munzner, 2014

Categorical vs. Ordered

- **Categorical** data has no implicit ordering, but might have a hierarchical structure

- **Ordered** data has implicit ordering

Example: Cars Dataset

Name	Economy	Cylinders	Displacement	Horsepower	Weight	Acceleration
Mazda RX4	21	6	160	110	2,62	16,46
Mazda RX4 Wag	21	6	160	110	2,875	17,02
Datsun 710	22,8	4	108	93	2,32	18,61
Hornet 4 Drive	21,4	6	258	110	3,215	19,44
Hornet Sportabout	18,7	8	360	175	3,44	17,02
Valiant	18,1	6	225	105	3,46	20,22
Duster 360	14,3	8	360	245	3,57	15,84
Merc 240D	24,4	4	146,7	62	3,19	20
Merc 230	22,8	4	140,8	95	3,15	22,9
Merc 280	19,2	6	167,6	123	3,44	18,3
Merc 280C	17,8	6	167,6	123	3,44	18,9
Merc 450SE	16,4	8	275,8	180	4,07	17,4
Merc 450SL	17,3	8	275,8	180	3,73	17,6
Merc 450SLC	15,2	8	275,8	180	3,78	18
Cadillac	10,4	8	472	205	5,25	17,98

Exercise: Forest Cover Type dataset

Attribute	**Values**	**Type**
Elevation	Meters	
Wilderness area	Boolean	
Slope	Degrees	
Soil type	Boolean	
Cover type	Integer [1,7]	

Ordering Direction

- Sequential

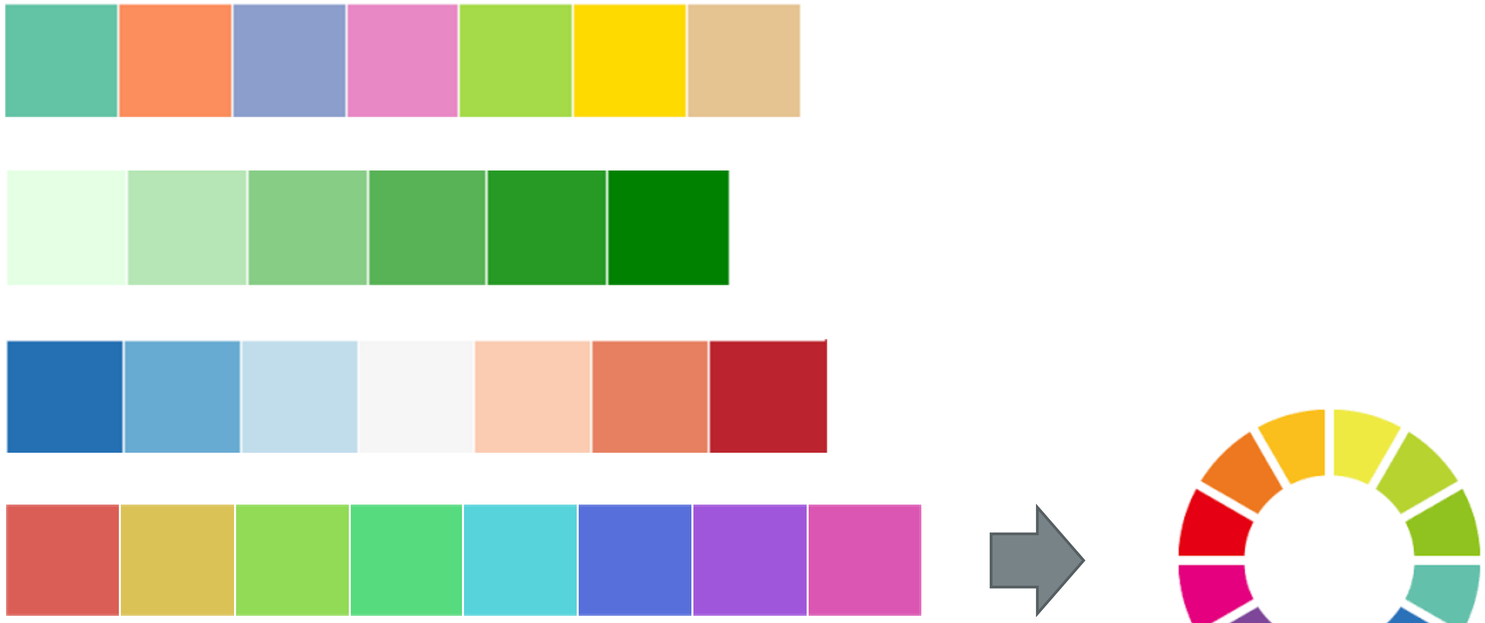
- Diverging

- Cyclic

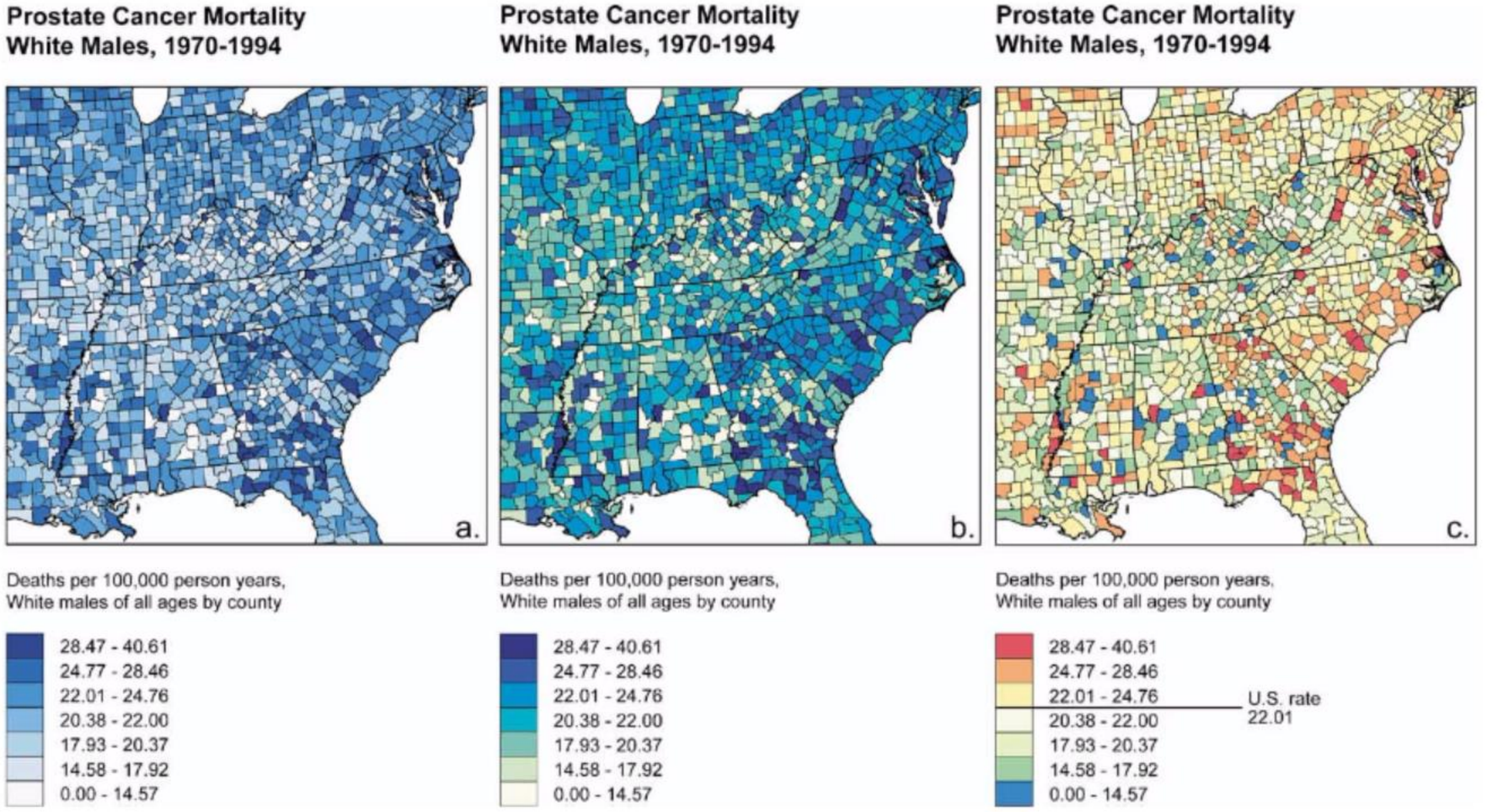
Munzner, 2014

Colors

- Categorical
- Sequential
- Diverging
- Cyclic

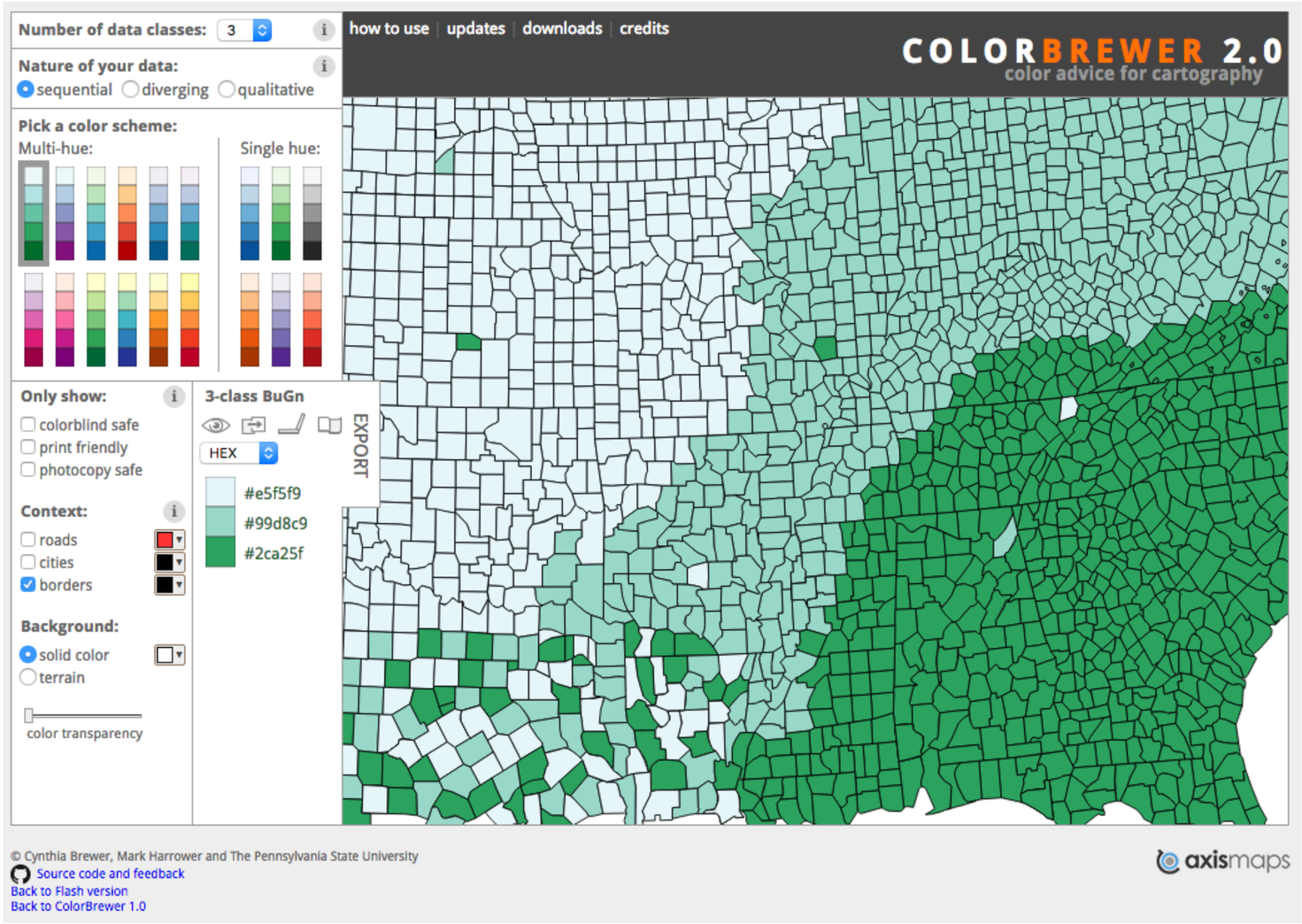


https://stanford.edu/~mwaskom/software/seaborn/tutorial/color_palettes.html



C. A. Brewer, "Basic Mapping Principles for Visualizing Cancer Data Using Geographic Information Systems (GIS)," American Journal of Preventive Medicine, vol. 30, no. 2, Supplement, pp. S25–S36, Feb. 2006.

Example: ColorBrewer2.org



Dataset Availability

- Adding/deleting items
- Changing a value of an existing item

Munzner, 2014

Semantics

The real world **meaning** of the data.

→ Keys **vs.** values

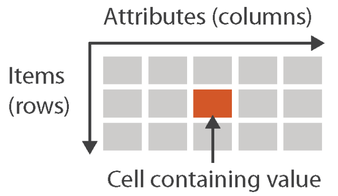
→ Spatial and continuous **vs.** non-spatial and discrete

→ Temporal **vs.** time-varying

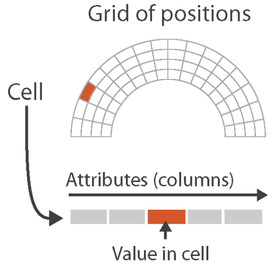
Keys vs. Values

- Key is an index that is used to look up the value.
  - Implicit key – index
  - Explicit key – attribute
- Keys must be unique for each item.

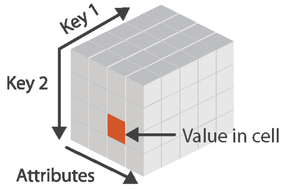
➔ Tables



➔ Fields (Continuous)



➔ *Multidimensional Table*



Munzner, 2014

Which types are suitable as a key **in a table**?

✓ → Categorical

→ Ordered

✓ → *Ordinal*

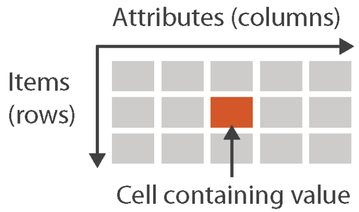
✗ → *Quantitative*

Munzner, 2014

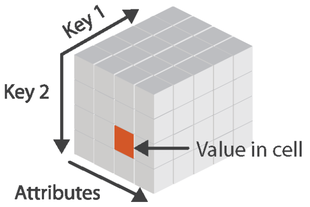
Keys and values in tables

- Determining which attributes are keys and which are values may be a part of the visualization process.
- Two dimensional (table) → Multidimensional.

➔ Tables



➔ *Multidimensional Table*

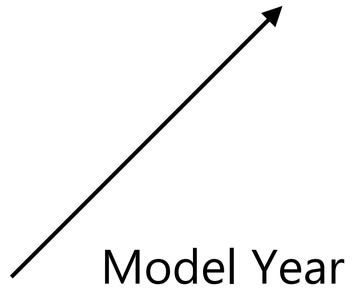


Munzner, 2014

Name	Economy	Cylinders	Displacement	Horsepower	Weight	Acceleration
Mazda RX4	21	6	160	110	2,62	16,46
Mazda RX4 Wag	21	6	160	110	2,875	17,02
Datsun 710	22,8	4	108	93	2,32	18,61
Hornet 4 Drive	21,4	6	258	110	3,215	19,44
Hornet Sportabout	18,7	8	360	175	3,44	17,02
Valiant	18,1	6	225	105	3,46	20,22
Duster 360	14,3	8	360	245	3,57	15,84
Merc 240D	24,4	4	146,7	62	3,19	20
Merc 230	22,8	4	140,8	95	3,15	22,9
Merc 280	19,2	6	167,6	123	3,44	18,3
Merc 280C	17,8	6	167,6	123	3,44	18,9
Merc 450SE	16,4	8	275,8	180	4,07	17,4
Merc 450SL	17,3	8	275,8	180	3,73	17,6
Merc 450SLC	15,2	8	275,8	180	3,78	18
Cadillac Fleetwood	10,4	8	472	205	5,25	17,98
Lincoln Continental	10,4	8	460	215	5,424	17,82
Chrysler Imperial	14,7	8	440	230	5,345	17,42
Fiat 128	32,4	4	78,7	66	2,2	19,47
Honda Civic	30,4	4	75,7	52	1,615	18,52
Toyota Corolla	33,9	4	71,1	65	1,835	19,9
Toyota Corona	21,5	4	120,1	97	2,465	20,01
Dodge Challenger	15,5	8	318	150	3,52	16,87
AMC Javelin	15,2	8	304	150	3,435	17,3
Camaro Z28	13,3	8	350	245	3,84	15,41
Pontiac Firebird	19,2	8	400	175	3,845	17,05
Fiat X1-9	27,3	4	79	66	1,935	18,9
Porsche 914-2	26	4	120,3	91	2,14	16,7
Lotus Europa	30,4	4	95,1	113	1,513	16,9
Ford Pantera L	15,8	8	351	264	3,17	14,5
Ferrari Dino	19,7	6	145	175	2,77	15,5
Maserati Bora	15	8	301	335	3,57	14,6
Volvo 142E	21,4	4	121	109	2,78	18,6

Multidimensional Tables

Name	Economy	Cylinders	Displacement	Horsepower	Weight	Acceleration
Mazda RX4	21	6	160	110	2,62	16,46
Mazda RX4 Wag	21	6	160	110	2,875	17,02
Datsun 710	22,8	4	108	93	2,32	18,61
Hornet 4 Drive	21,4	6	258	110	3,215	19,44
Hornet Sportabout	18,7	8	360	175	3,44	17,02
Valiant	18,1	6	225	105	3,46	20,22



Multidimensional Tables

Name	Economy	Cylinders	Displacement	Horsepower	Weight	Acceleration	Year
Mazda RX4	21	6	160	110	2,62	16,46	
Mazda RX4 Wag	21	6	160	110	2,875	17,02	
Datsun 710	22,8	4	108	93	2,32	18,61	
Hornet 4 Drive	21,4	6	258	110	3,215	19,44	
Hornet Sportabout	18,7	8	360	175	3,44	17,02	
Valiant	18,1	6	225	105	3,46	20,22	

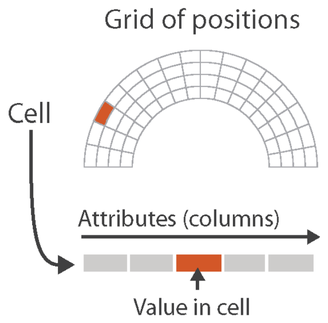
Multidimensional Tables

Name	Economy	Cylinders	Displacement	Horsepower	Weight	Acceleration	Year
Mazda RX4	21	6	160	110	2,62	16,46	1977
Mazda RX4	22	6	160	115	2,5	16,7	1978
Mazda RX4 Wag	21	6	160	110	2,875	17,02	1977
Mazda RX4 Wag	21	6	165	110	2,875	17,02	1978
Datsun 710	22,8	4	108	100	2,32	18,61	1977
Datsun 710	22,8	4	108	93	2,32	18,61	1978
Hornet 4 Drive	21,4	6	258	110	3,215	19,44	1977
Hornet 4 Drive	21,4	6	258	110	5	19,44	1978
Hornet Sportabout	18,7	8	360	175	3,44	17,02	1977
Hornet Sportabout	18,7	8	360	175	3,44	20	1978
Valiant	18,1	6	225	105	3,46	20,22	1977
Valiant	18,1	8	225	105	3,46	20,22	1978

Keys and values in fields

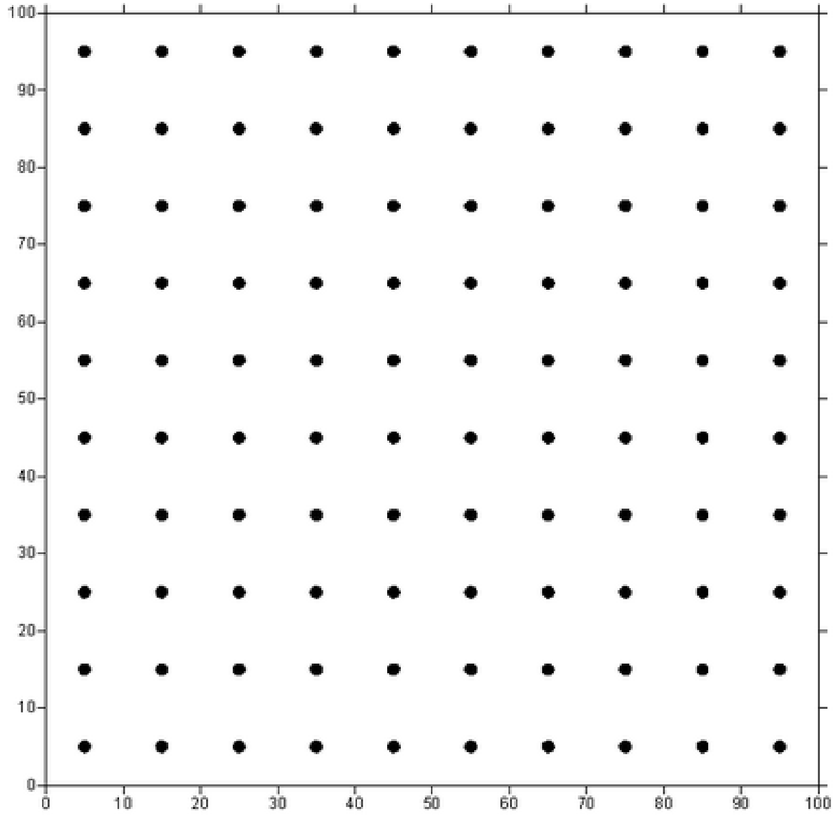
- Independent/dependent variables (synonyms)
- Values can be estimated throughout the sampled range (not only for sampled points) via interpolation
- Keys: spatial locations, time, etc.

➔ Fields (Continuous)



Munzner, 2014

Position is **key**



www.spatialanalysisonline.com

Which type could be used as a key for **a field**?

✗ → Categorical

→ Ordered

✗ → *Ordinal*

✓ → *Quantitative*

Munzner, 2014

Temporal Attributes

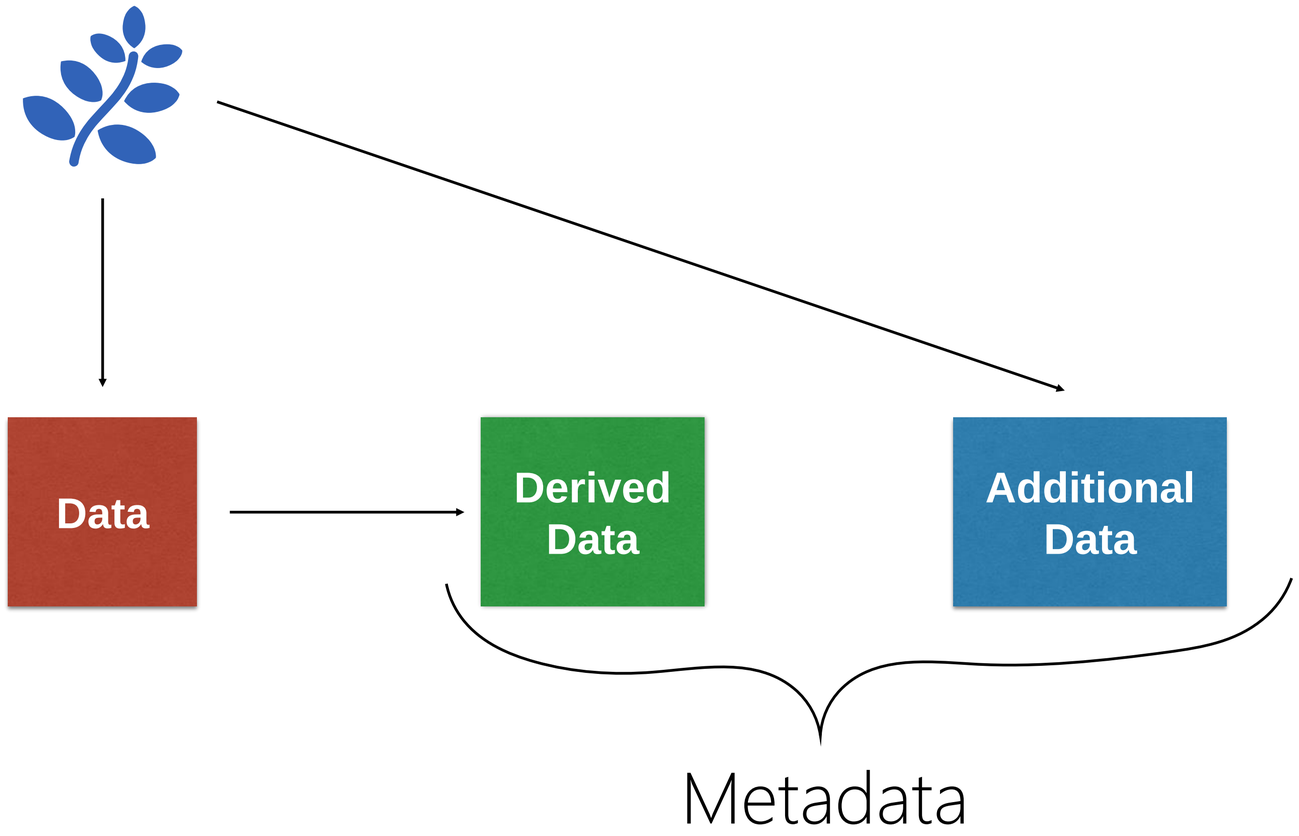
- Temporal information relates to time.
- Time has hierarchy and periodicity.
- Hierarchy is not strict (weeks do not fit cleanly into months).
- Temporal attribute can be a value (elapsed time) or key (date).

Time-varying Data

- Time is one of the **key** attributes
  - e.g., location of an animal measured every second.
- Time-series data
  - ordered (by time) sequence of time-value pairs.

Metadata

Metadata



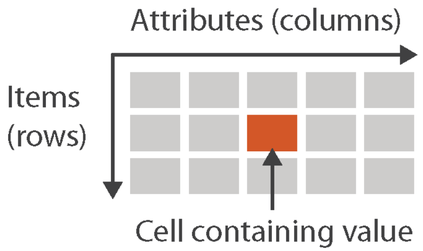
Metadata

- Secondary data (not of the first interest)
- Metadata is additional data that gives information and describes the data
  - Could be derived from the data
- From visualization point of view is the same as data

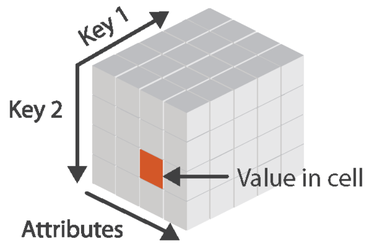
Biological Data

Exercise: What type is it?

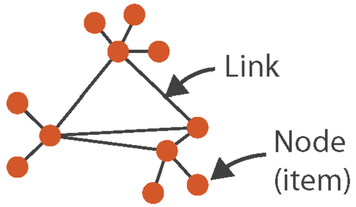
➔ Tables



➔ *Multidimensional Table*

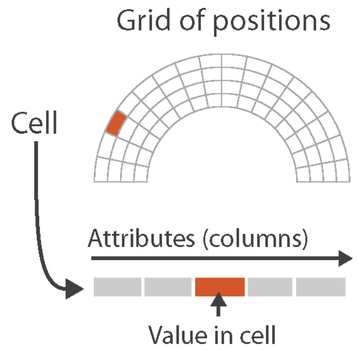


➔ Networks

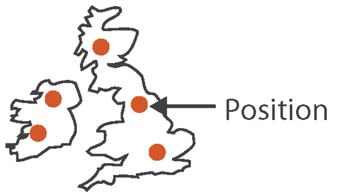


➔ *Trees*

➔ Fields (Continuous)



➔ Geometry (Spatial)



➔ **Sets**

➔ **Lists**

➔ **Clusters**

➔ **etc.**

Munzner, 2014

Exercise: What attributes and ordering does it have?

- **Attribute Types**
  - Categorical

  - Ordered
    - *Ordinal*

    - *Quantitative*

- **Ordering Direction**
  - Sequential

  - Diverging

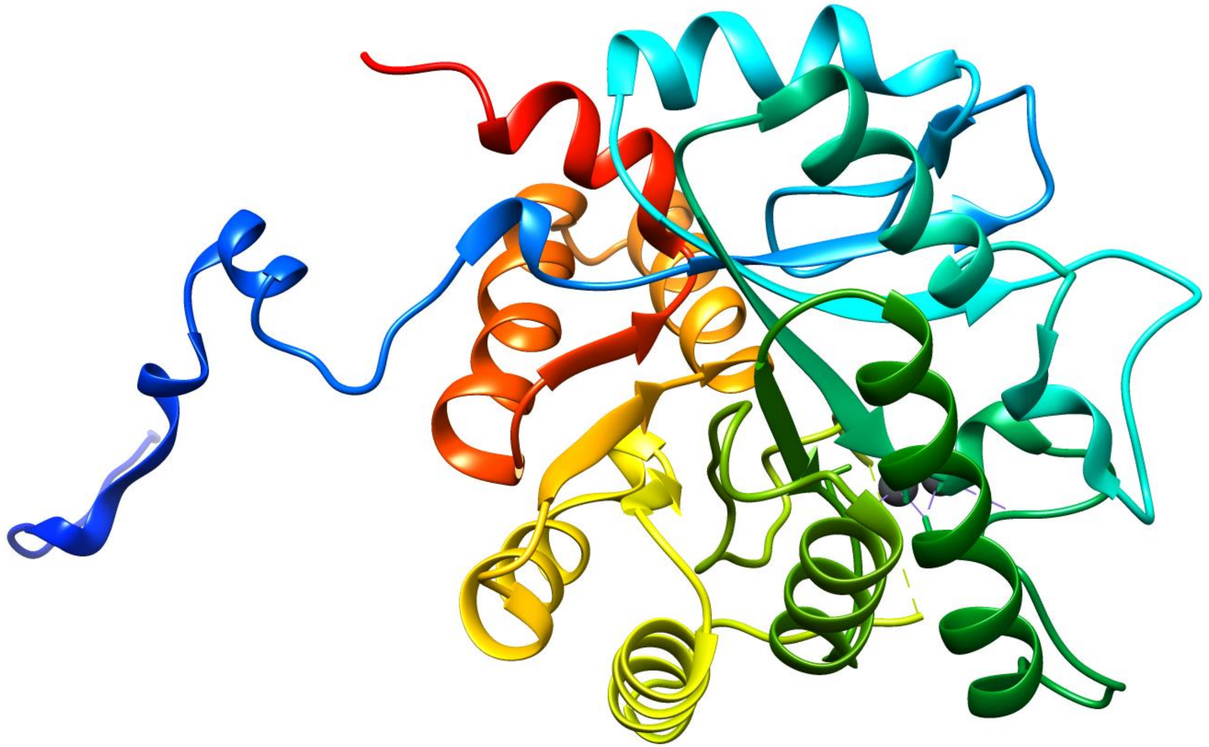
  - Cyclic

Munzner, 2014

Sequences

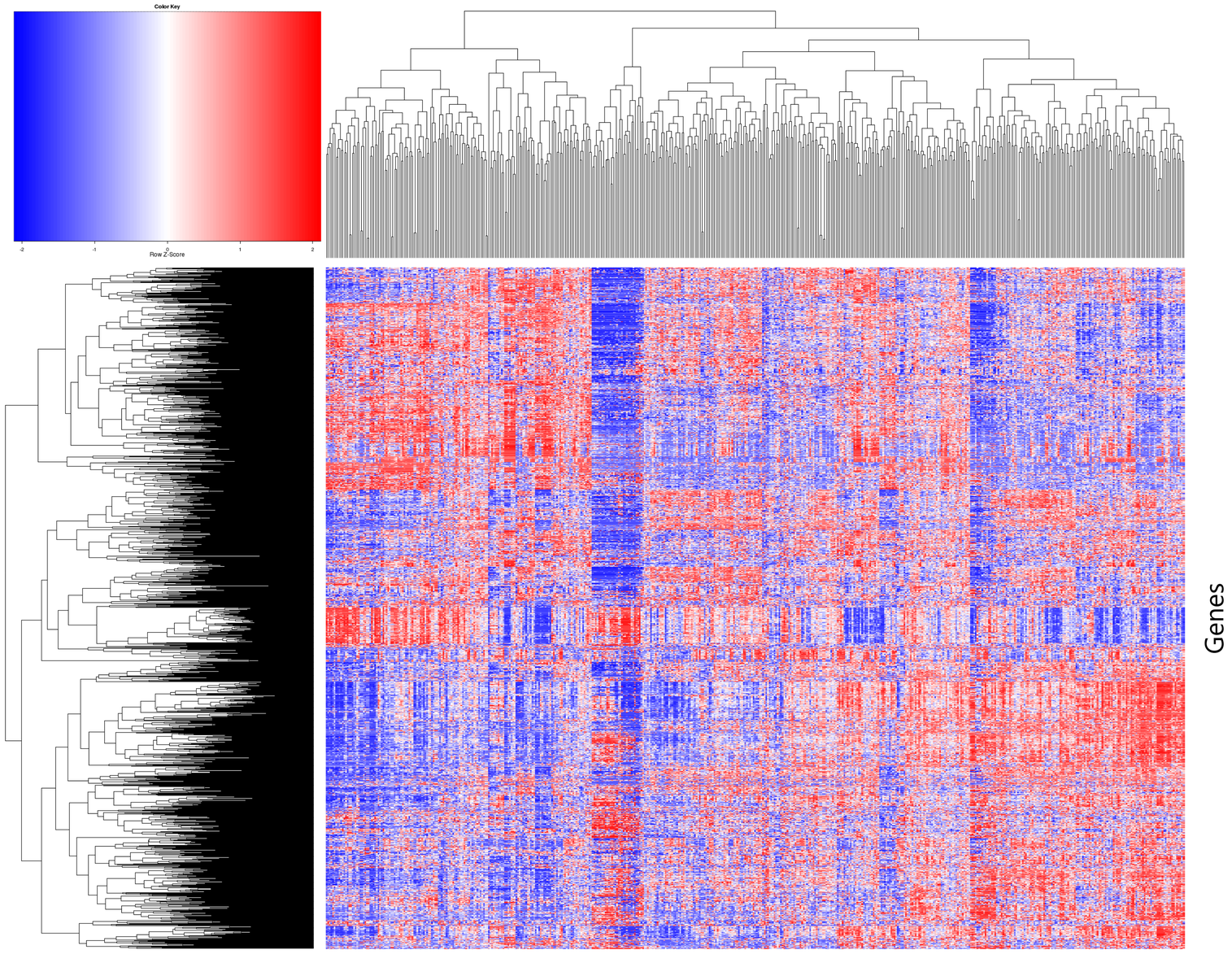
RREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR

Macromolecules

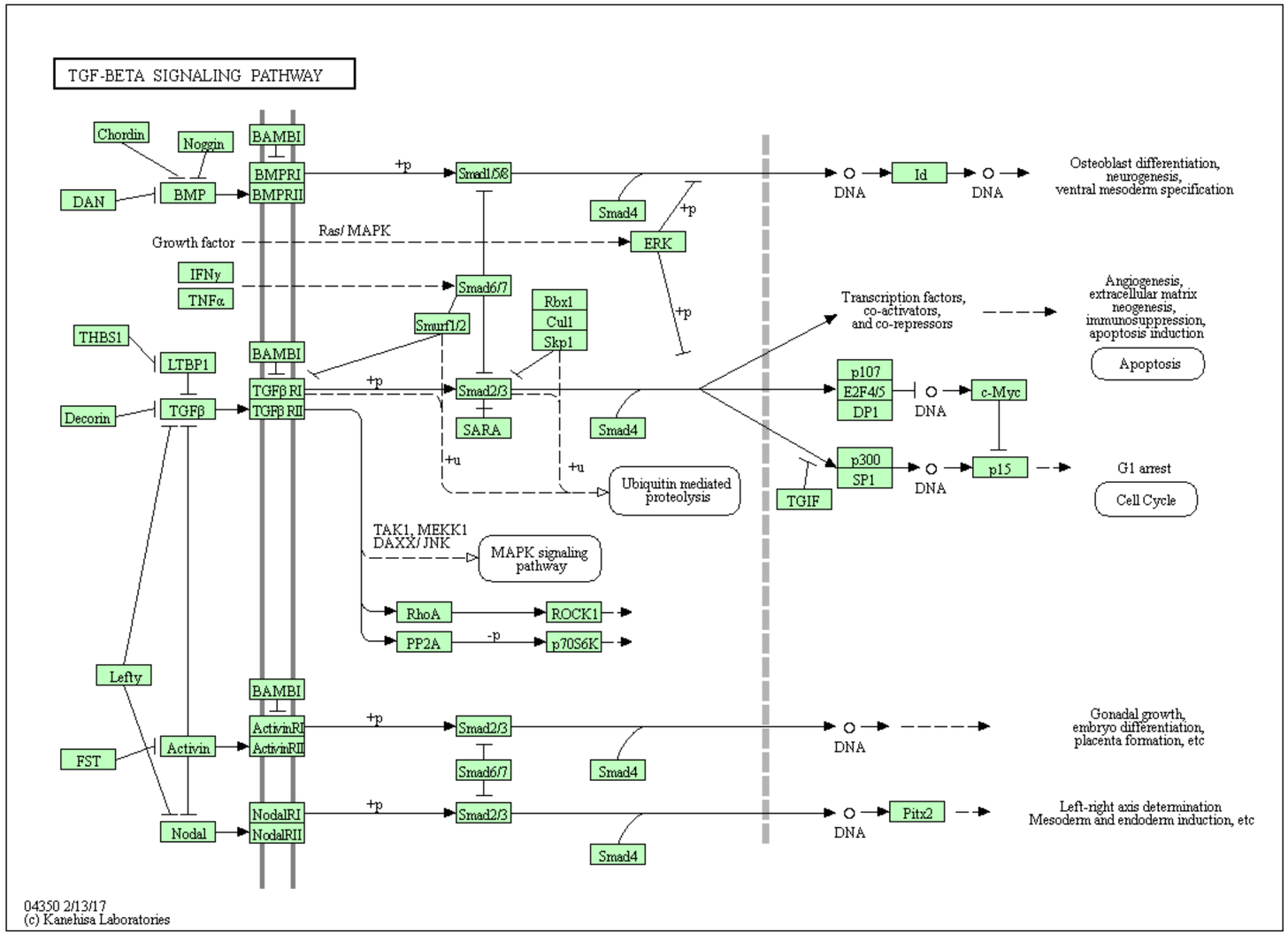


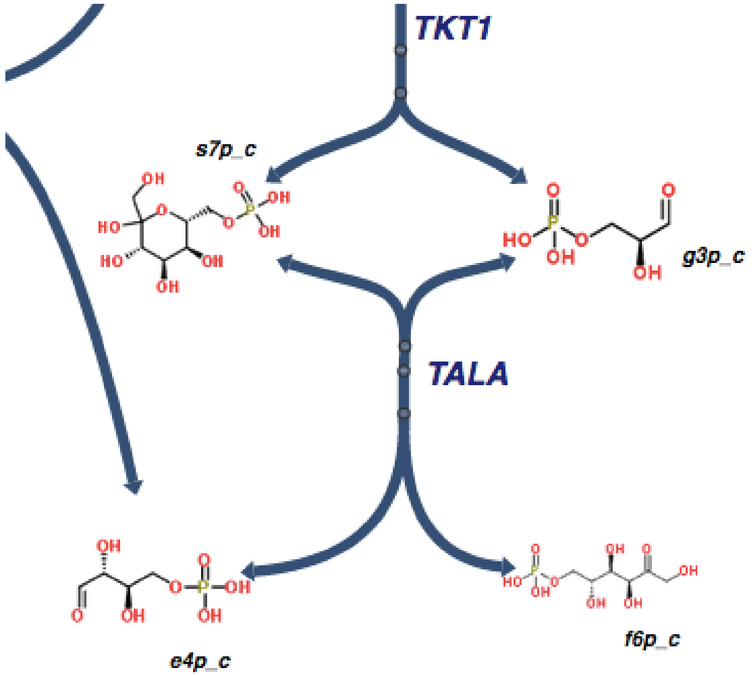
RREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKR

Gene Expression



Pathways


TGF-BETA SIGNALING PATHWAY
Chordin
Noggin
BAMBI
BMPRI
BMPRII
BMP
DAN
+p
Smad1/5/8
Smad4
DNA
Id
DNA
Osteoblast differentiation,
neurogenesis,
ventral mesoderm specification
Growth factor
Ras/ MAPK
ERK
+p
IFNγ
TNFα
Smad6/7
Rbx1
Cul1
Skp1
Smurf1/2
Transcription factors,
co-activators,
and co-repressors
Angiogenesis,
extracellular matrix
neogenesis,
immunosuppression,
apoptosis induction
Apoptosis
THBS1
LTBP1
BAMBI
TGFβ RI
TGFβ RII
TGFβ
Decorin
+p
Smad2/3
SARA
Smad4
p107
E2F4/5
DP1
DNA
c-Myc
p300
SP1
DNA
p15
G1 arrest
Cell Cycle
TGIF
+u
+u
Ubiquitin mediated
proteolysis
TAK1, MEKK1
DAXX/ JNK
MAPK signaling
pathway
RhoA
ROCK1
PP2A
-p
p70S6K
Lefty
BAMBI
ActivinRI
ActivinRII
Activin
FST
+p
Smad2/3
Smad6/7
Smad4
DNA
Gonadal growth,
embryo differentiation,
placenta formation, etc
NodalRI
NodalRII
Nodal
+p
Smad2/3
Smad4
DNA
Pitx2
Left-right axis determination
Mesoderm and endoderm induction, etc
04350 2/13/17
(c) Kanehisa Laboratories


TKT1
s7p_c
g3p_c
TALA
e4p_c
f6p_c

Biological Data

- Sequences
- Alignments
- Motifs
- Protein function
- Macromolecules
- Secondary structure
- Protein surface
- Expression levels
- Gene interactions
- Gene regulation data
- Pathways
- Phylogenies
- Metagenomics data:
  - Single sample
  - Sampled from a field
- Scientific literature

To sum up

What?

Datasets

Data Types

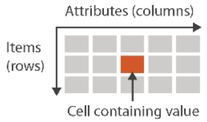
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

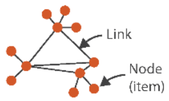
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

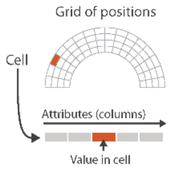
- Tables



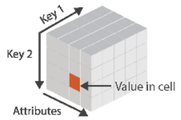
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic

Munzner, 2014

Next Lecture:

Why: Task Abstraction