Data(set) Types and Semantics

Torsten Möller

Schedule

Week	Date	Tuesday	Friday
1	Mar 05/08	No Lecture	Introduction (TM) pdf [Munzner Ch. 1] Rosling at TED 2006; Krulwich at RadioLab
2	Mar 12/15	Rektorstag	Tableau Q&A-A1 (YJ) Tutorial
3	Mar 19/22	Design Principles (TM) pdf	Data(set) Types and Semantics (TM) pdf Task Typology
	Mar 19	Due: A1 (23:55)	
4	Mar 26/29	Easter	Easter
	Mar 31	Due: A2 (23:55)	
5	Apr 02/05	Easter	Easter
6	Apr 09/12	Tasks (FW) pdf [Munzner Ch. 2+3] Task Typology	No Lecture
7	Apr 16/19	Visual Encoding Principles (TM) pdf [Munzner Ch. 5+10] Learning perceptual kernels for vis design Livingstone: What Art can tell us about the brain (Vis 2008 keynote);	Tableau Tutorial (YJ)

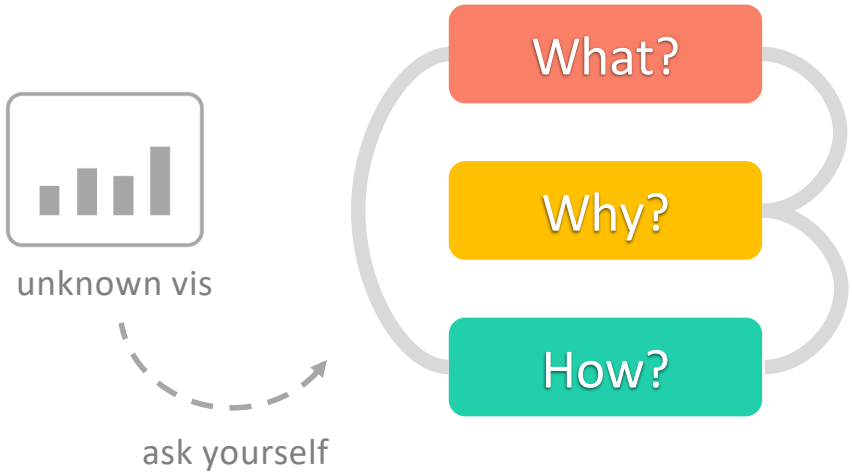
- Last week: Design Principles
  - Graphical Integrity
  - Design Principles
  - Design Elements
- Today: Data(set) Types and Semantics
  - Semantics vs. Types
  - Data(set) & Attribute Types
  - Attribute + Dataset Semantics
  - Derived Data
- Follow-up:
  - Visual Encoding
  - Tasks

Reading

- Munzner, “Visualization Analysis and Design”: Chapter 2 (What: Data Abstraction)
- Shneiderman, “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations,” IEEE Symposium on Visual Languages, 1996
- Heer+Shneiderman, “Interactive Dynamics for Visual Analysis,” Communications of the ACM 2012.
- Amar et al., “Low-level components of analytic activity in information visualization,” InfoVis 2005.
- Brehmer+Munzer, “A Multi-Level Typology of Abstract Visualization Tasks,” InfoVis 2013.

Three-part analysis framework

to analyze any existing visualization



What data is shown?

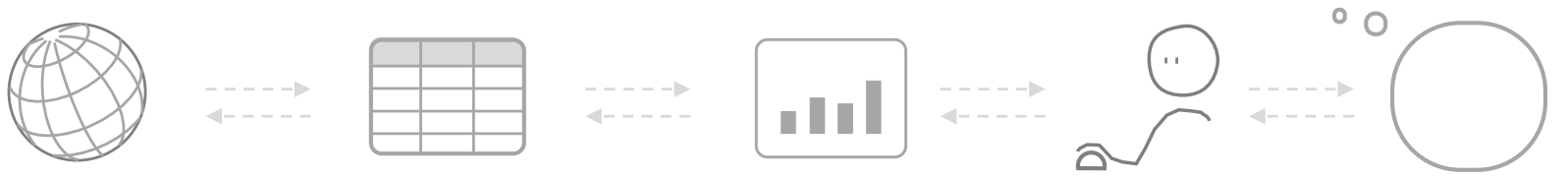
Why is the task being performed?

How is the visualization constructed?

Munzner, Visualization Analysis and Design, 2014, p. 17 ff.

Three-part analysis framework

to analyze any existing visualization



world data(set) vis user task

What? How? Why?

Munzner, Visualization Analysis and Design, 2014, p. 17 ff.

Semantics vs. Types

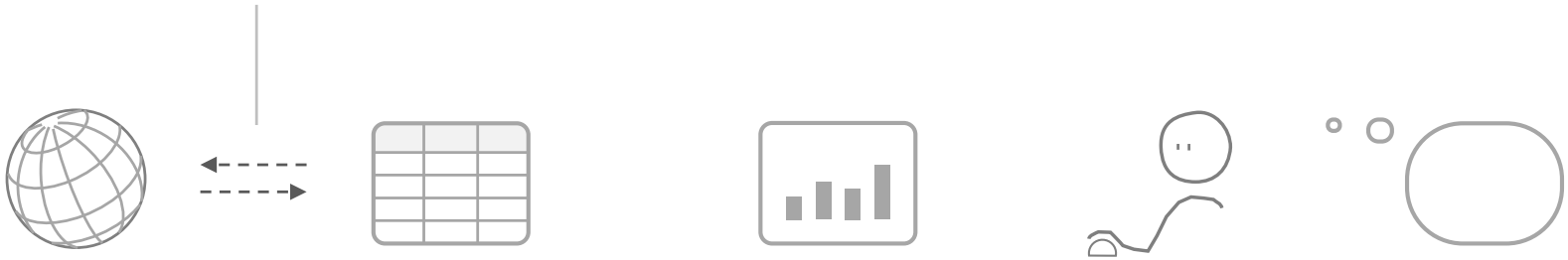
Semantics vs. Type

- Semantics: real-world meaning of data
- Type: abstract classification with implications on
  - mathematical operations
  - data structure (how to store)
- Given semantics – type will follow

Semantics vs. Type

Munzner, Visualization Analysis and Design, 2014, p. 17 ff.

semantics of data = its real-world meaning



world data(sets) vis user task

different types = its structural / mathematical interpretation

Data (attribute) Types

Data Types

Twitter User
ScaredOfTheDark(net)
@RetweetableN14

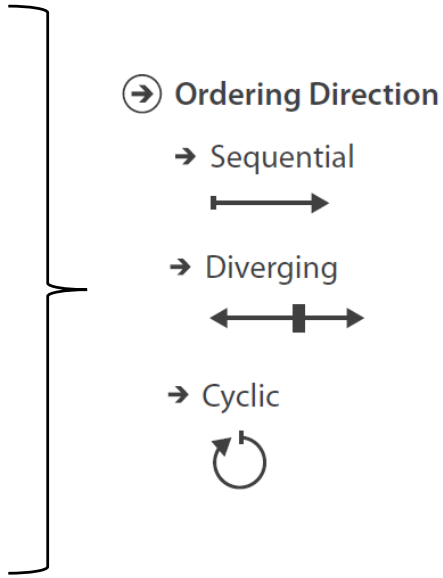
Data Types

- Items
  Individual entity
- Attributes
  Property that is measured, observed
- Links
  Express relationship between items
- Positions
  Spatial data location (2D, 3D)
- Grids
  Sampling strategy for continuous data

Data (Attribute) Types

- Ordered
  - Quantitative
    10 inches, 23 cm, etc.

  - Ordinal
    Small, medium, large



- Categorical (Nominal, Qualitative)
  Apples, Oranges, Bananas,...

32	7/16/07	2-High	Medium Box	0.65	7/18/07

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
32	7/16/07	2-High	Medium Box	0.65	7/18/07

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

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

Record / 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

Dimension / 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	[illegible]	0.6	6/6/05
70	12/18/06	5-Low	[illegible]	0.59	12/23/06
70	12/18/06	5-Low	[illegible]	0.82	12/23/06
96	4/17/05	2-High	[illegible]	0.55	4/19/05
97	1/29/06	3-Medium	[illegible]	0.38	1/30/06
129	11/19/08	5-Low	[illegible]	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

1 = Quantitative

2 = Ordinal

3 = Nominal

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 Speci		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		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

1 = Quantitative

2 = Ordinal

3 = Nominal

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 Speci		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		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

1 = Quantitative

2 = Ordinal

3 = Nominal

Example Dataset: „World Happiness“

Description : Composed from multiple sources (summary)

- The **World Happiness Report** is a landmark survey of the state of global happiness, covering 155 countries in 2017. […] The scores are based on answers to a life evaluation question asked in the poll, where respondents rate the quality of their current lives on a scale from 1 to 10. (kaggle)

- The **CIA World Factbook** contains almanac-style information about the countries of the world, such as population, GDP, family income, internet access, cell phone subscriptions, etc. (link)

- The **Human Development Index** (HDI) (link)

- The **Inequality-adjusted Human Development Index** (INA) (link)

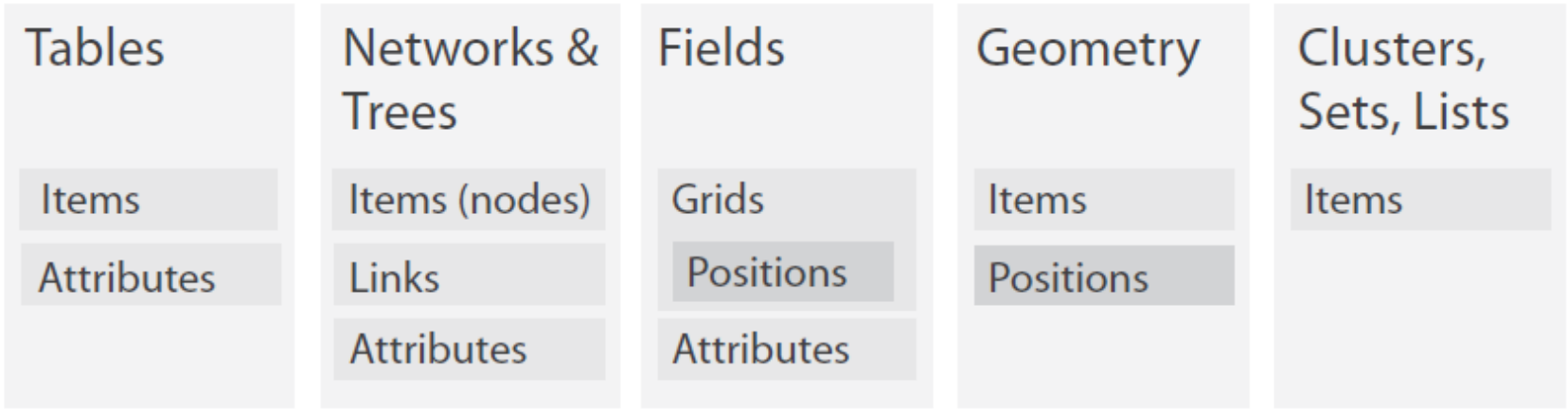
Dataset types

Dataset types

„A **dataset** is any collection of information that is the target of analysis.“

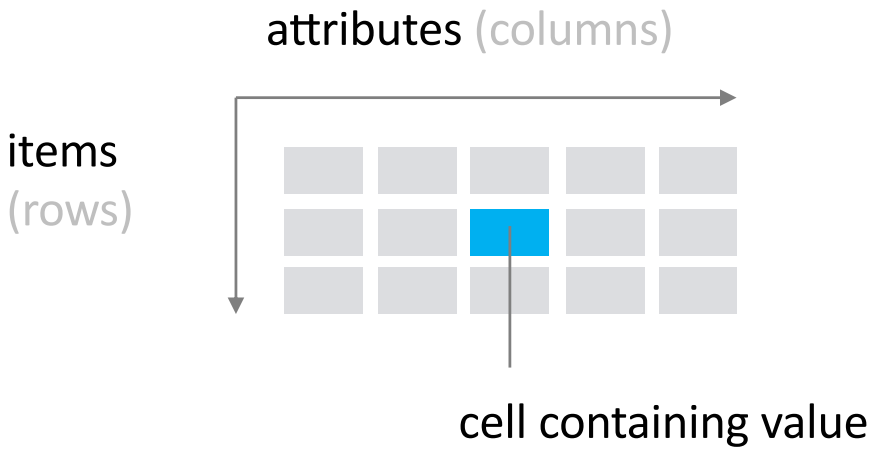
Munzner, Visualization Analysis and Design, 2014, p. 24

Basic types:



(+ e.g. Text + Log Files, …)

Tables



- Each data **item** in a new row
- Each column contains an **attribute**

Tableau Example

- Let‘s build our first visualization!

Tableau Example

- Dimensions - **Categorical**
- Measures - **Quantitative**

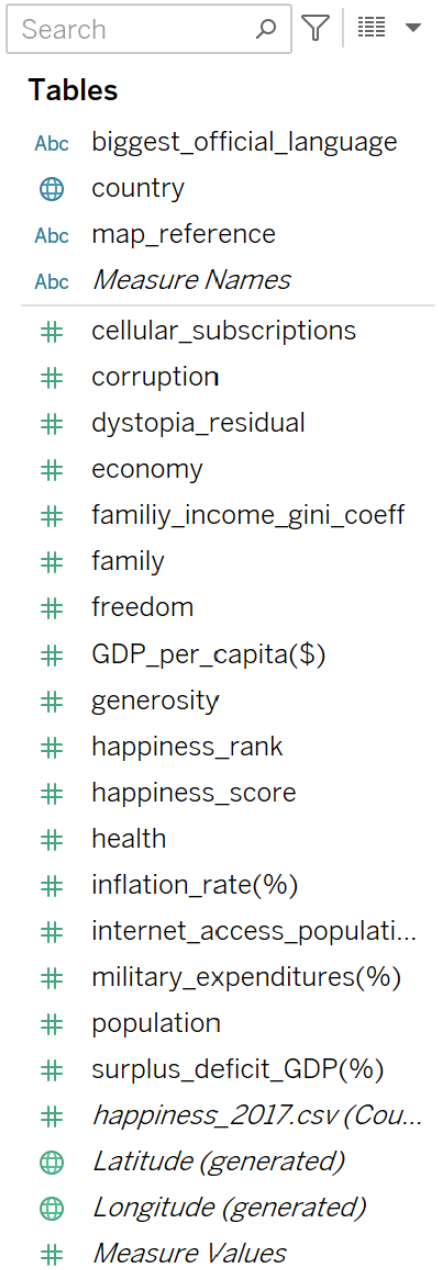


Tableau Example

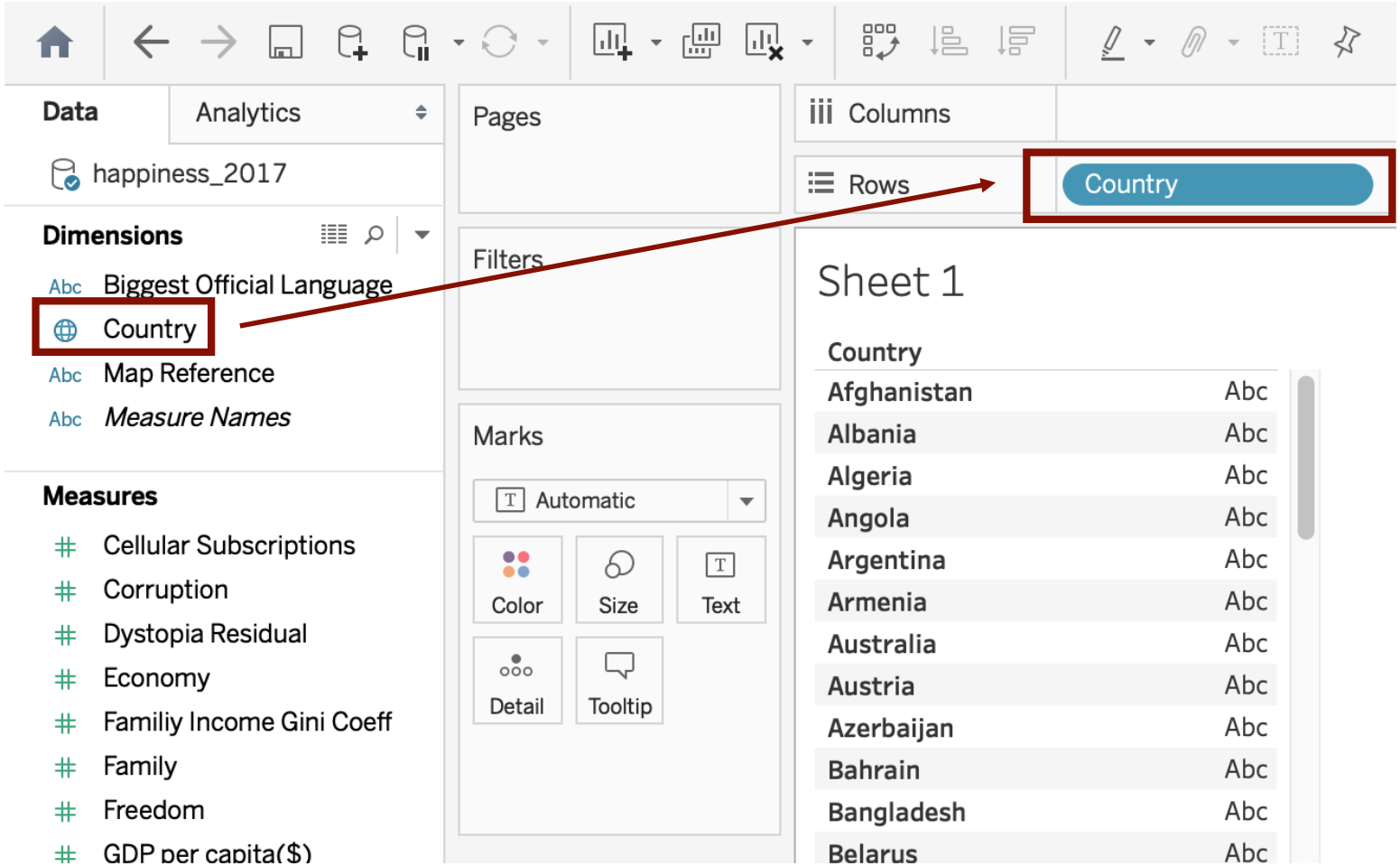


Tableau Example

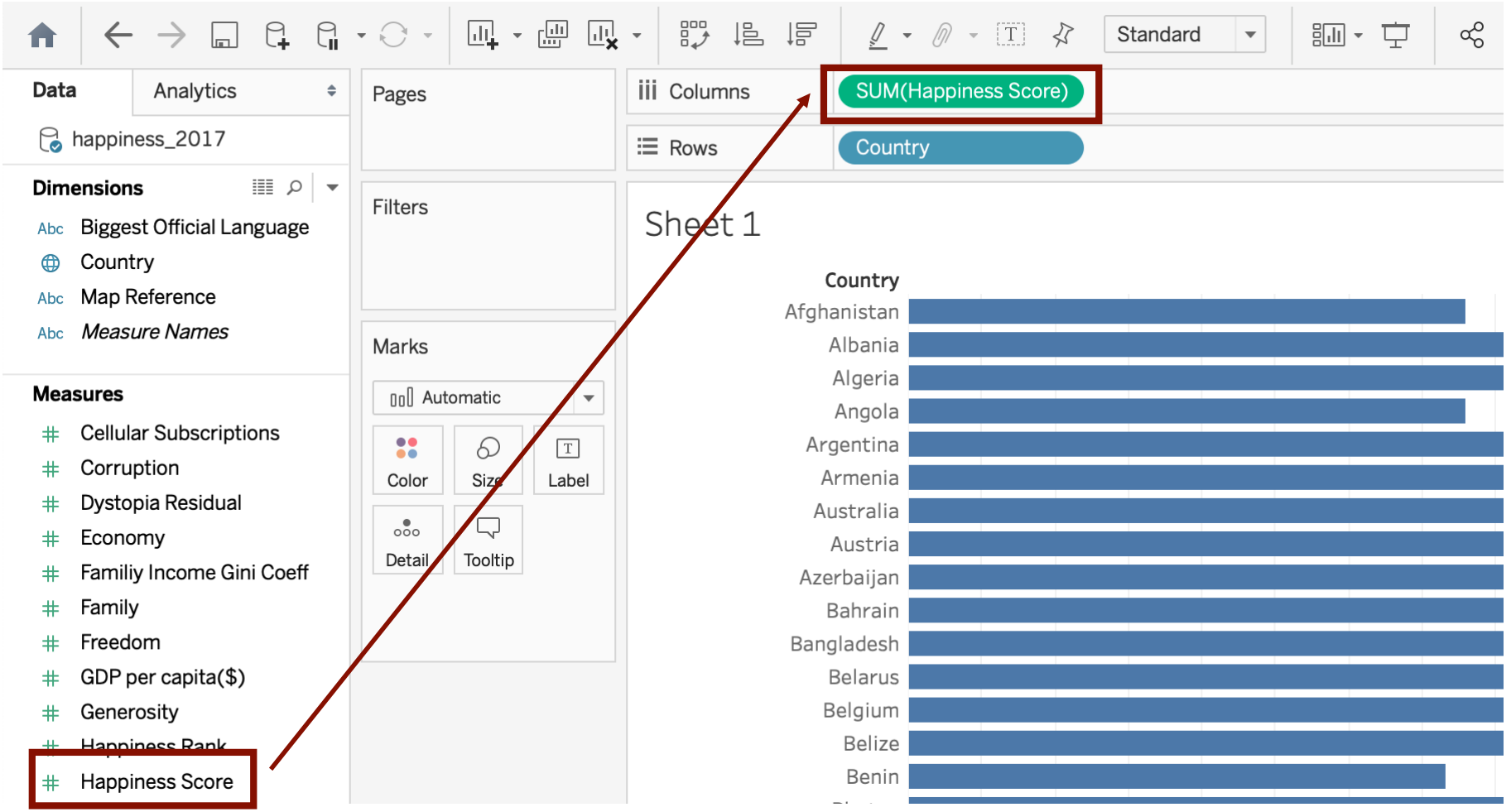


Tableau Example

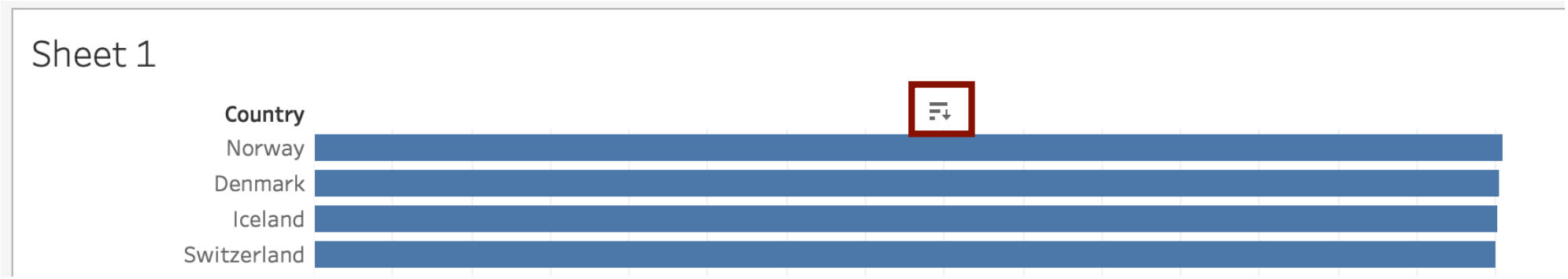


Tableau Example

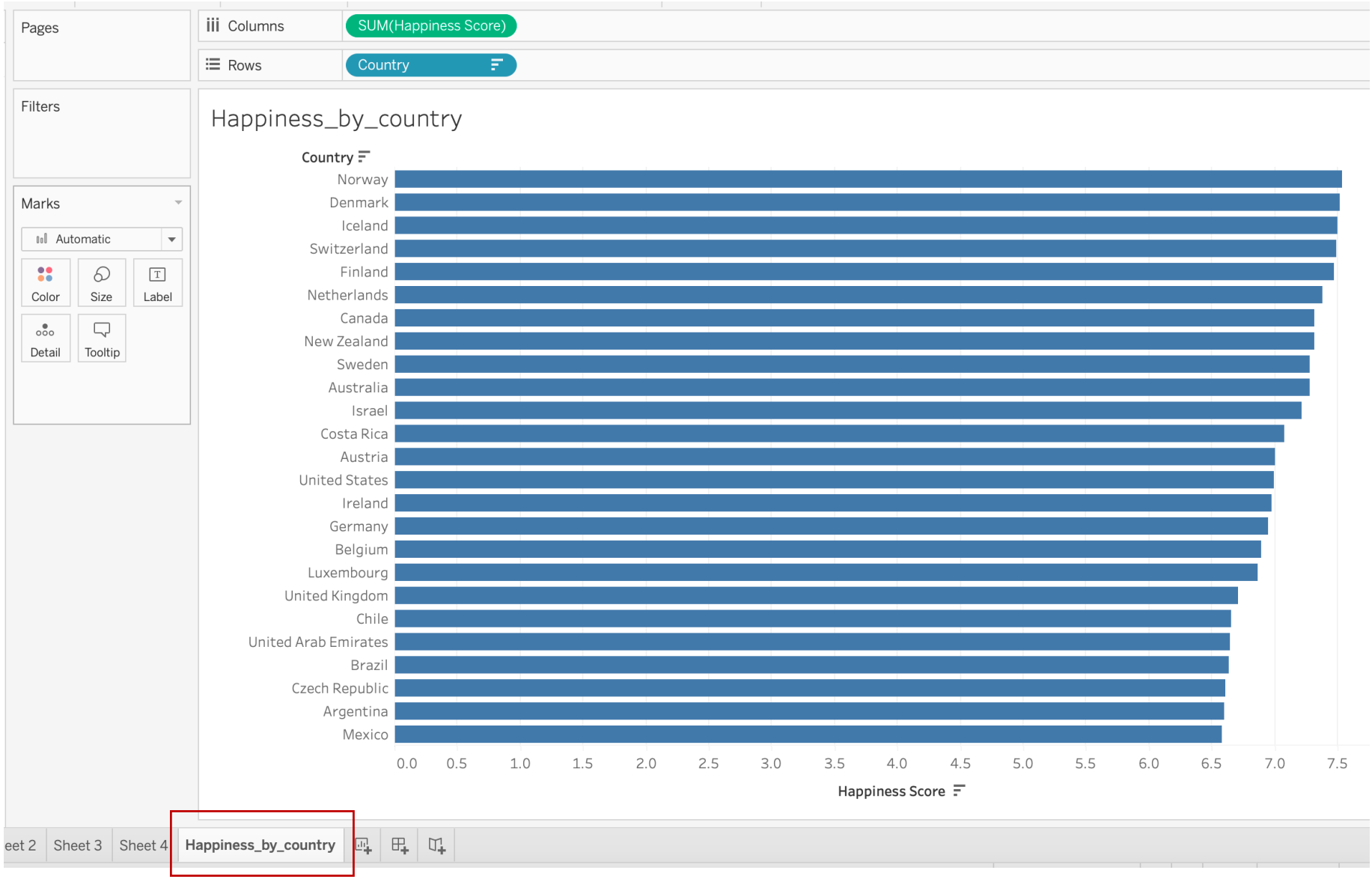


Tableau Example

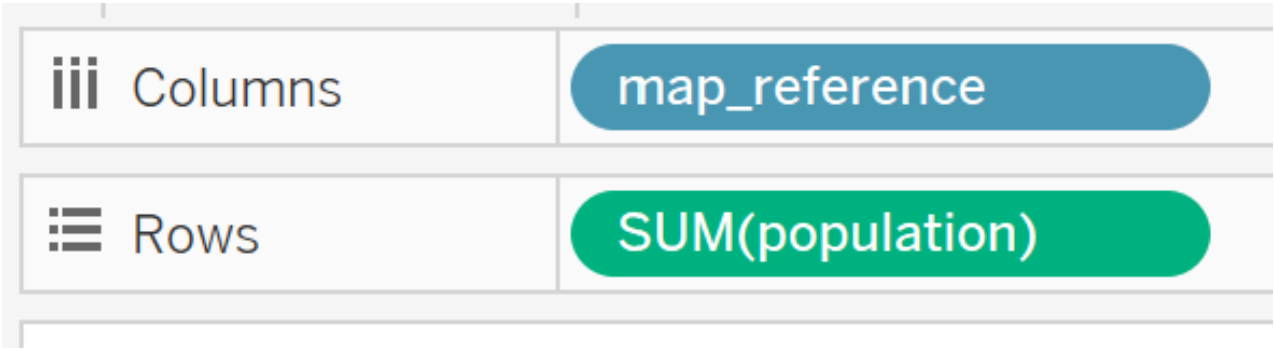


Tableau Example

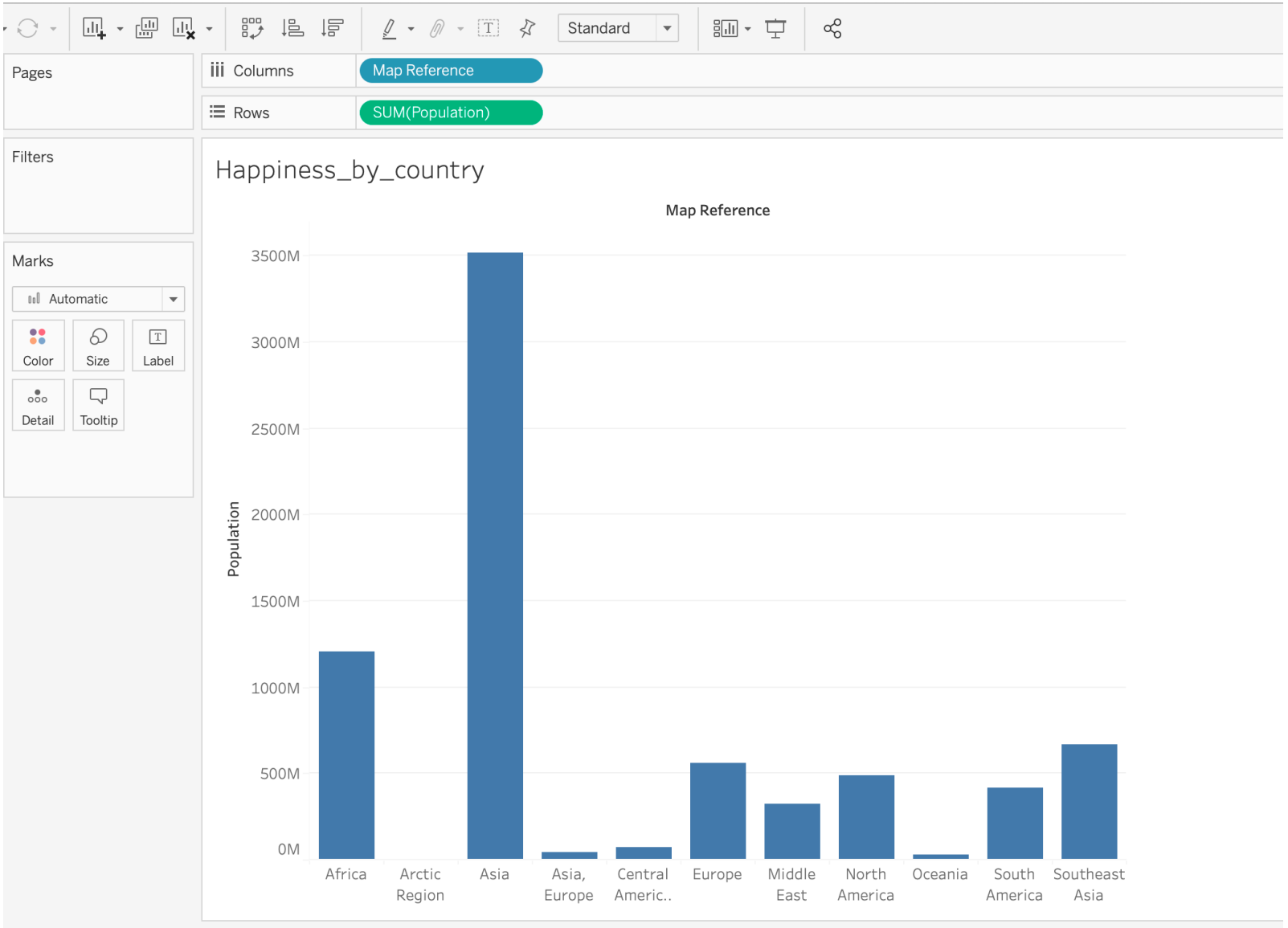


Tableau Example

What happens when we set
Population as a *Dimension*?

Tableau Example

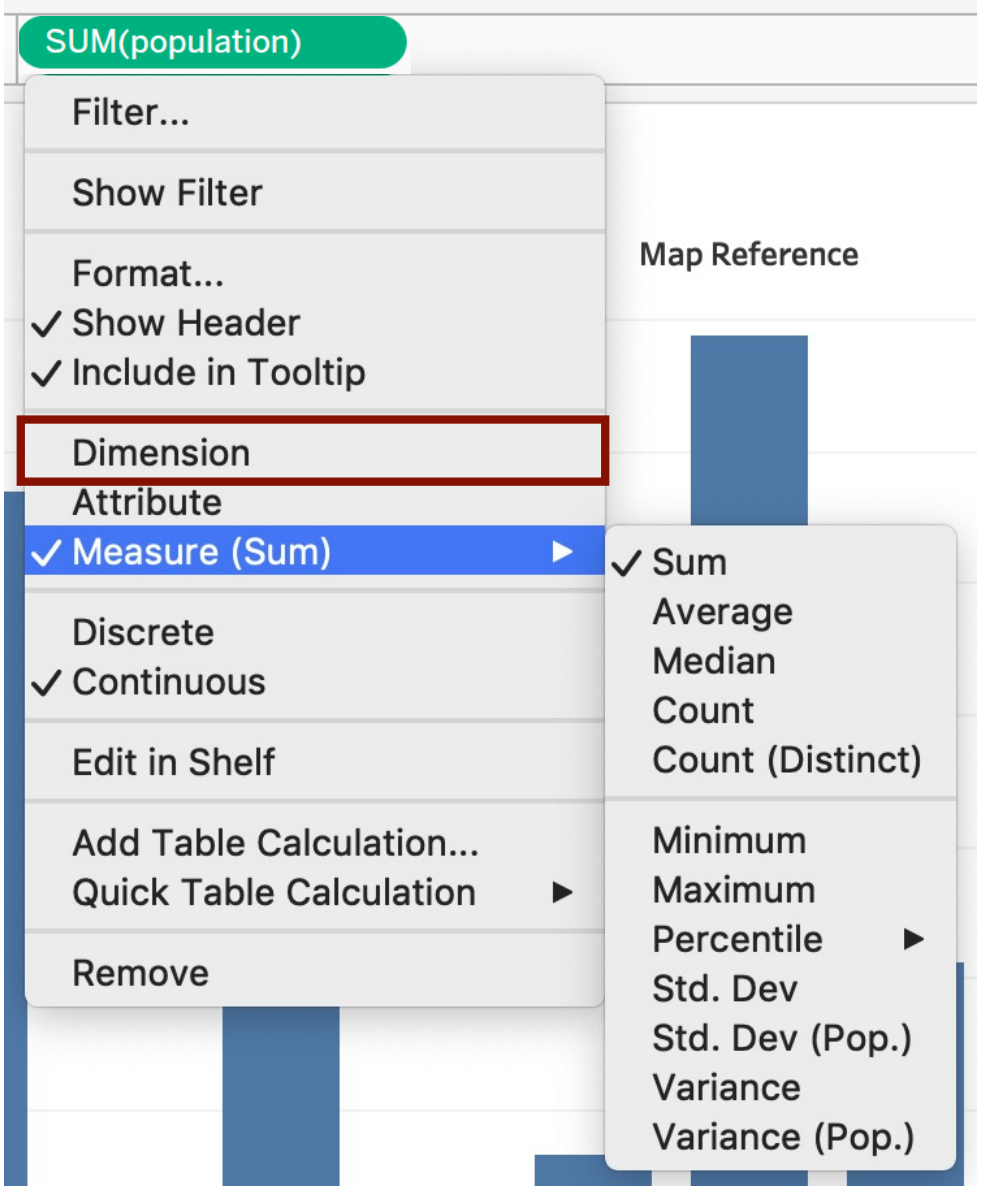


Tableau Example

Each bar is one country

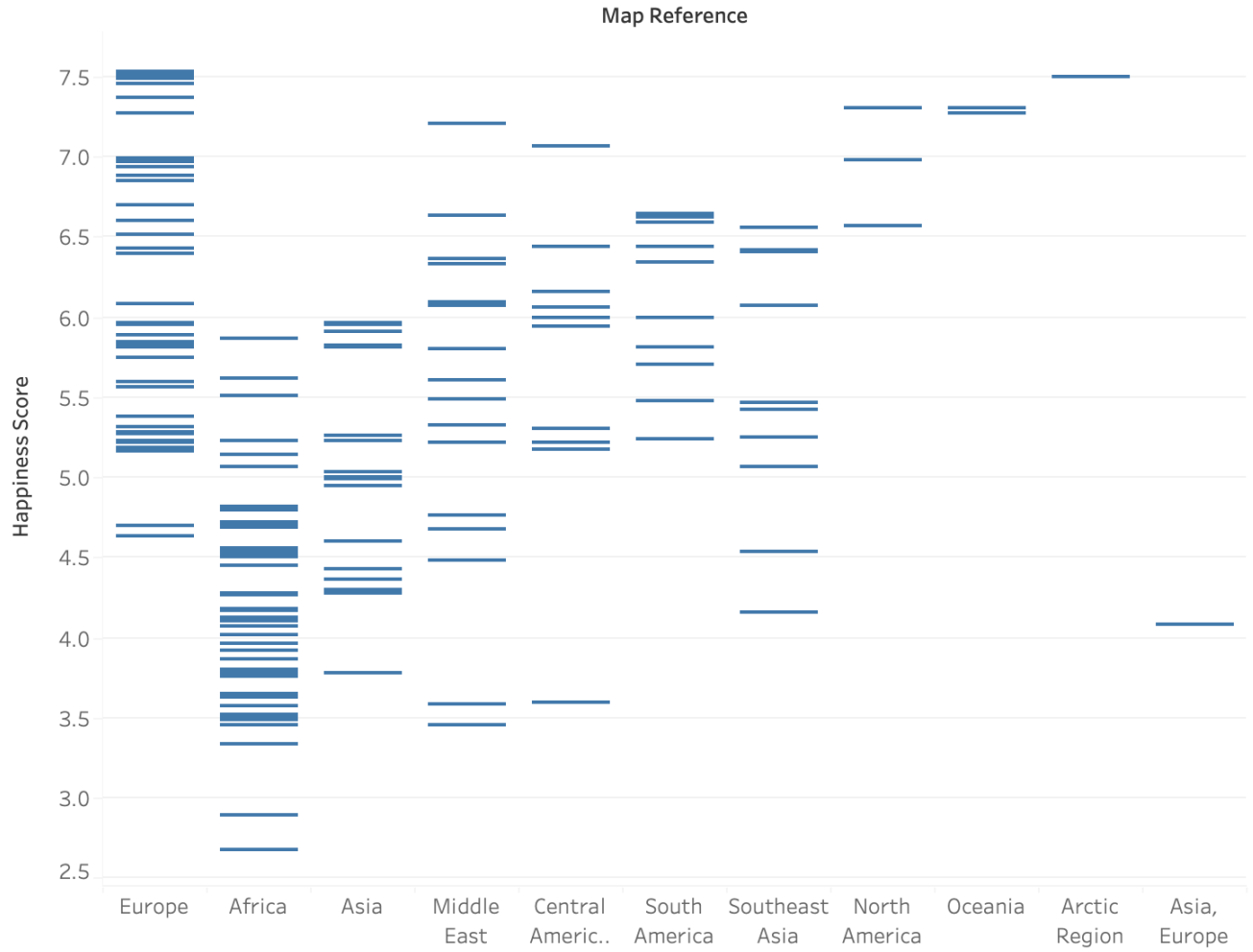


Tableau Example

Now tell me:

- Which Continent has the highest **total** Population?
- Which one has the highest **average** (per country)?

Tableau Example

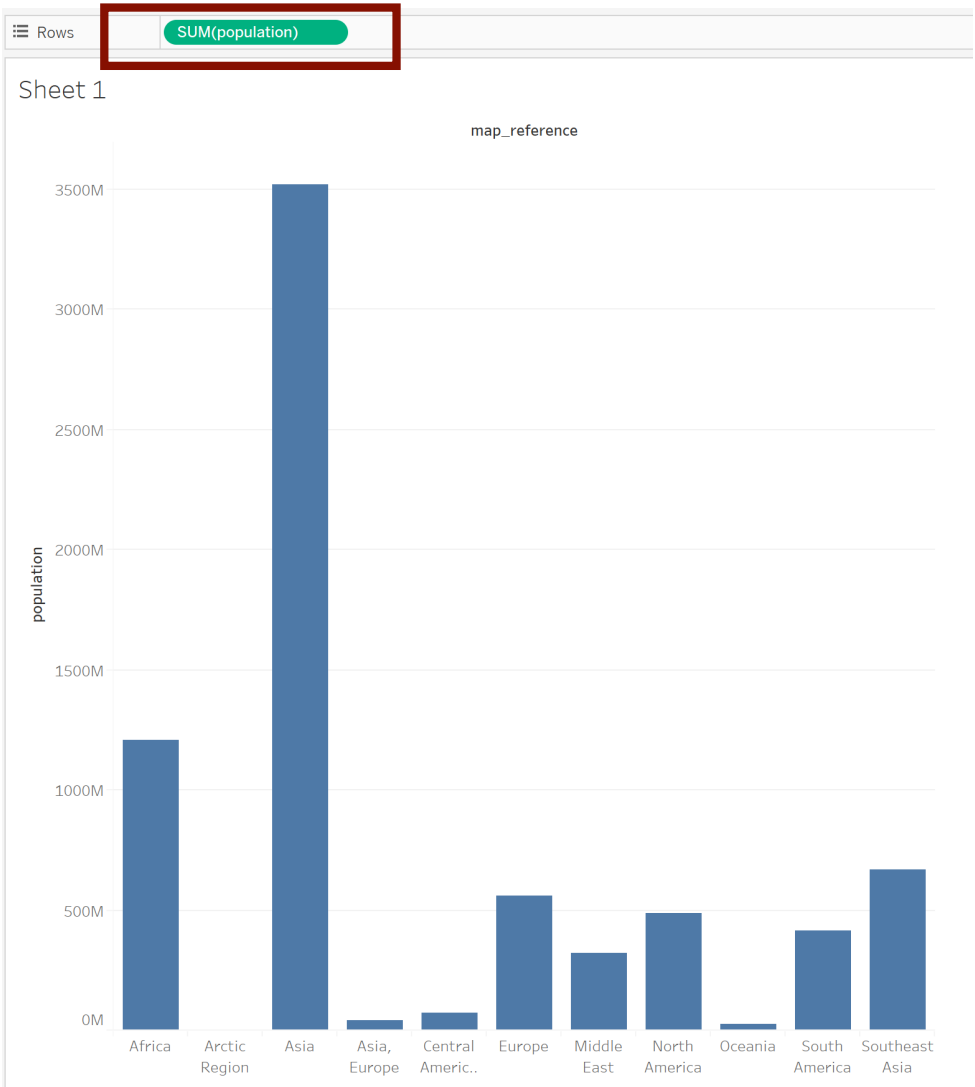


Tableau Example

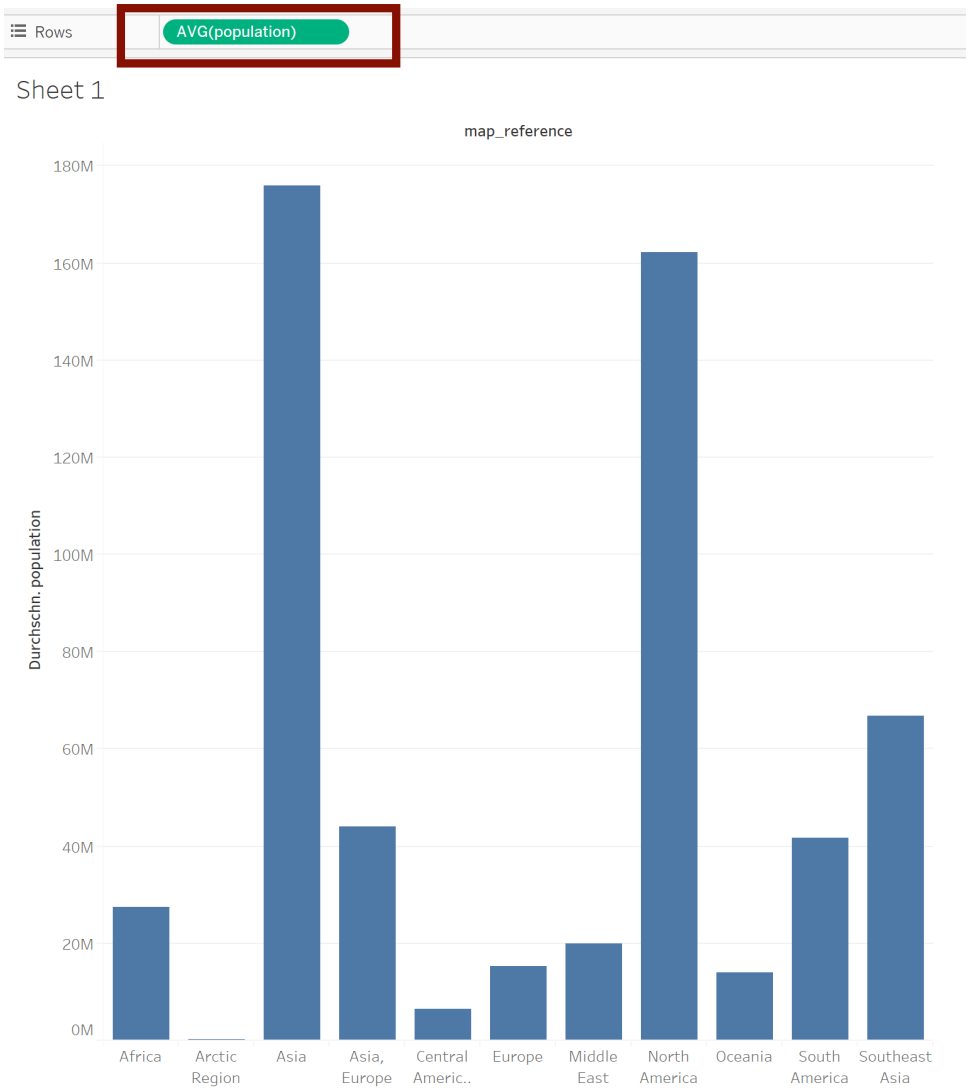
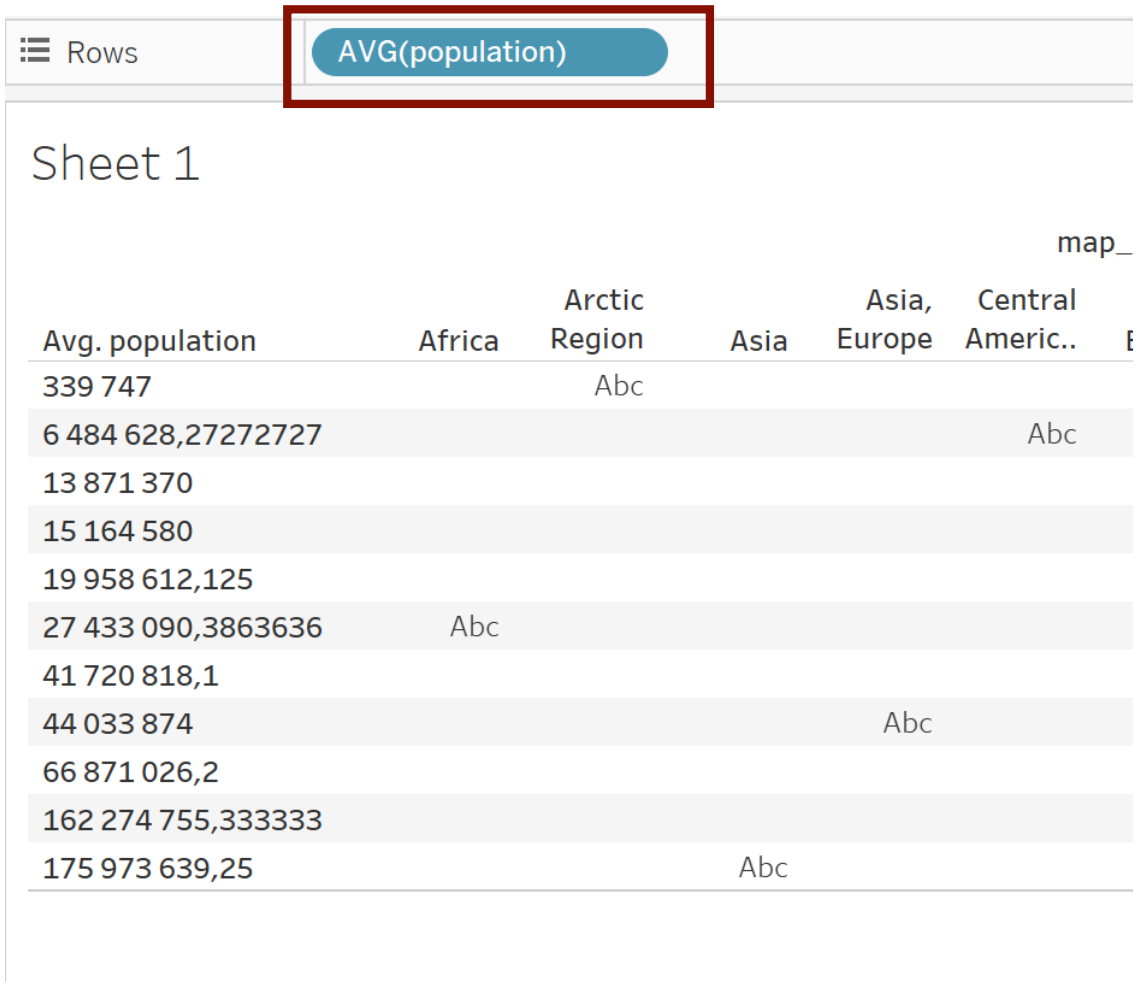


Tableau Example

What happens when we convert the **Population** to *Discrete*?

Tableau Example



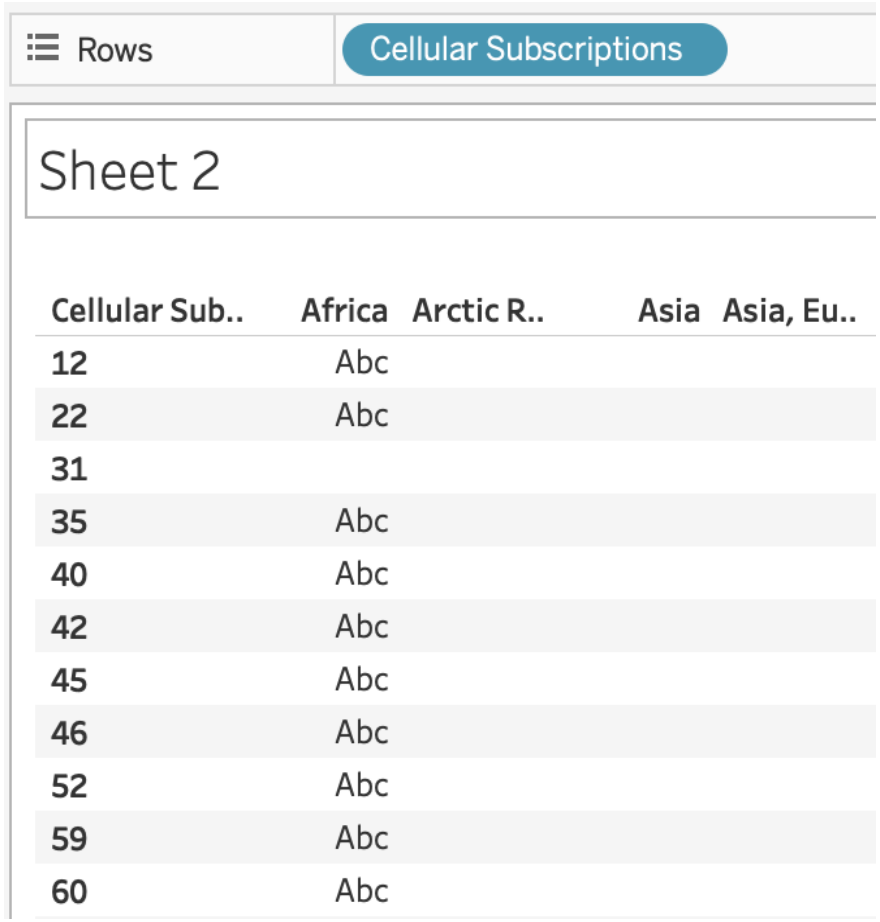
- 11 values are shown
- One for each Continent

Tableau Example

What happens when we convert **Population** to *Discrete* and to a *Dimension*?

Tableau Example

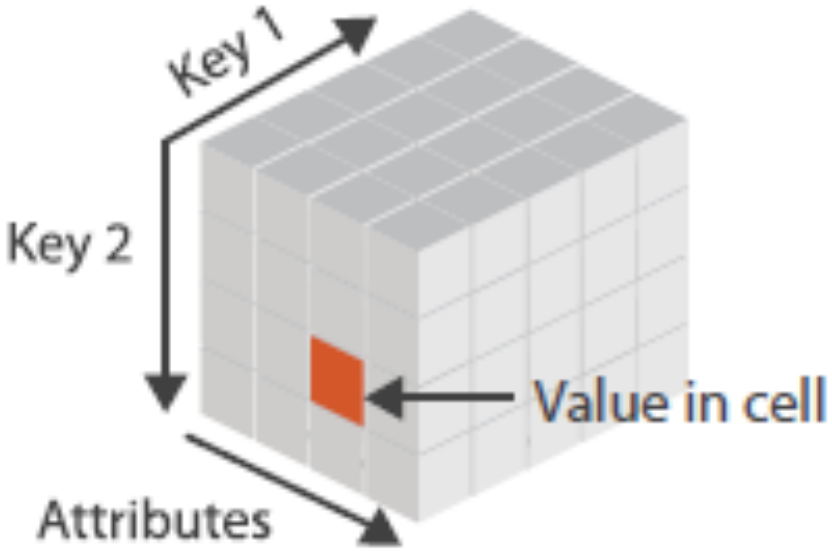
The values are shown at Country level



Cellular Sub..	Africa	Arctic R..	Asia	Asia, Eu..
12	Abc			
22	Abc			
31				
35	Abc			
40	Abc			
42	Abc			
45	Abc			
46	Abc			
52	Abc			
59	Abc			
60	Abc			

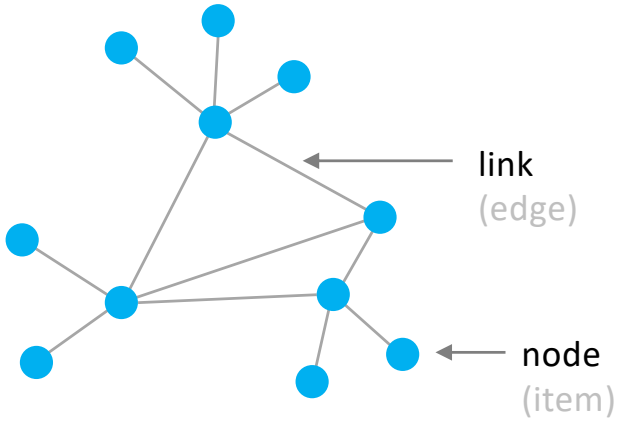
→ Lecture on Apr 23rd – Arrange Tables + Spatial Data

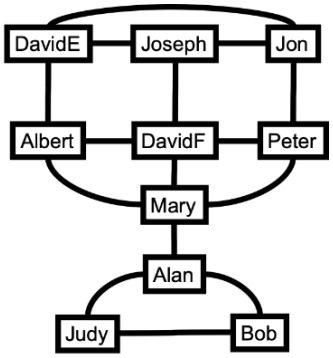
Multi-dimensional tables



Networks / Graphs

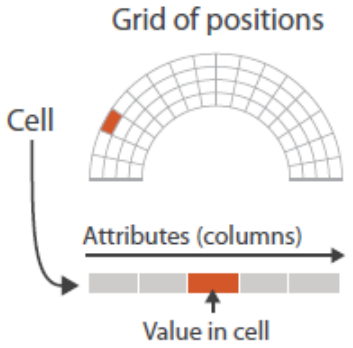
- Item = **node**
- **Link** between items = **edge**
- For example, social network:
  people + friendship ties
- Both links and nodes
  can have attributes
- Graphs can be represented by
  2 tables (node & edge list/table)
- **Trees** are strictly hierarchical graphs





→ Lecture on May 7th/14th – Arrange Networks/Trees

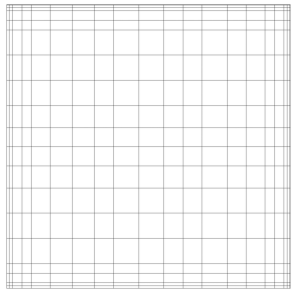
Spatial data

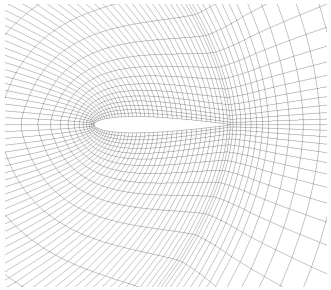


- Continuous dataset
- Specified through grids (connectivity) were data is sampled:

uniform rectilinear curvilinear/structured irregular

ParaView user guide





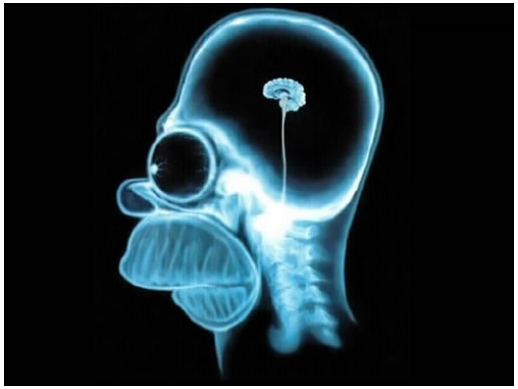
- Geometry

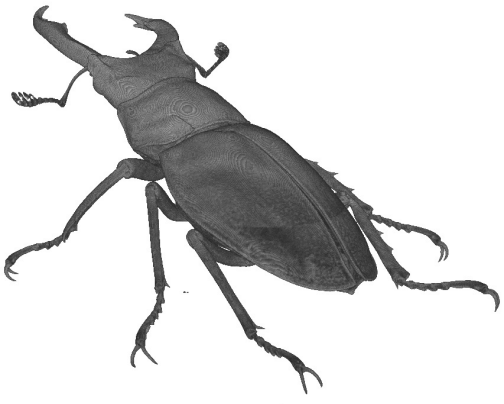
Spatial dimensions

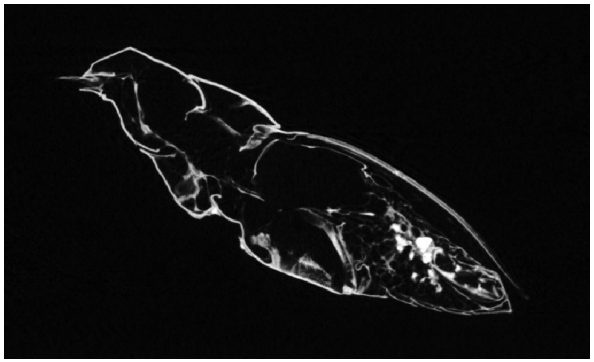
- 1D: refers to a single ‘length’ scale (e.g. height)
- 2D: geographical information
- 3D: medical / physics
- Time-varying:
  - 1D+time
  - 2D+time
  - 3D+time

Spatial values: Scalar data

- Mapping f: $R^n \rightarrow R$, $(x_1,...,x_n) \rightarrow y$
- n independent variables (keys) x_i (1D, 2D, or 3D, +time)
- Value y is just univariate
- Example:
  - 2D(/3D) grey-scale image data
  - MRI, CT,

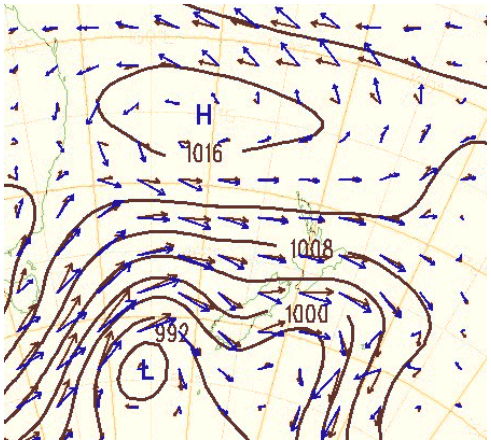






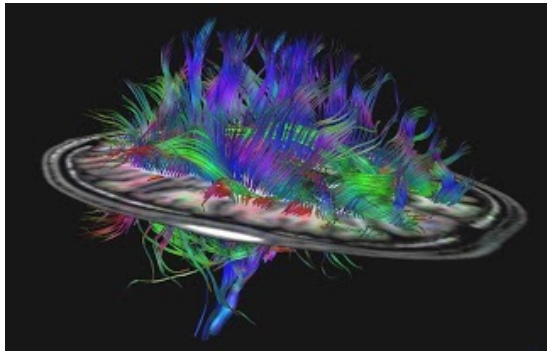
Spatial values: Vector data

- Mapping f: $R^n \rightarrow R^m$, $(x_1,...,x_n) \rightarrow (y_1,...,y_m)$
- Vector at each position represents direction and magnitude
- Usually, m=n
- Exceptions, e.g., due to projection
- Example:
  - Weather map (wind direction + speed)
  - Flow around airplane wings



Spatial values: Tensor data

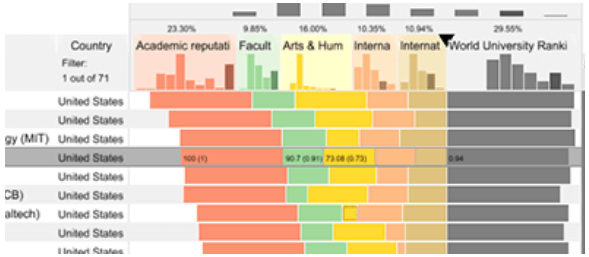
- Mapping f: $R^n \rightarrow R^m$, $(x_1,...,x_n) \rightarrow y_{i1,i2,...,ik}$
- Tensor of level k
  - Tensor of level 1 is a vector
  - Tensor of level 2 is a matrix
  - …
- Example:
  - Diffusion-tensor MRI
  - Stress-tensor (9 numbers representing forces acting in 3 orthogonal directions)



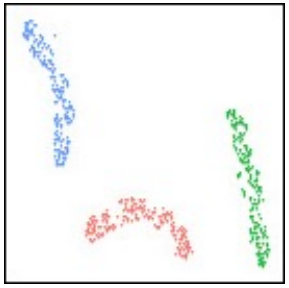
Collections: Clusters, Sets, Lists

How we group items

- Sets
  Unique items, unordered
- Lists
  Ordered, duplicates possible
- Clusters
  Groups of similar items



Gratzl et al., LineUp: Visual Analysis of Multi-Attribute Rankings, 2013



Abbas et al., ClustMe: A Visual Quality Measure for Ranking Monochrome Scatterplots based on Cluster Patterns, 2019

Text + Logs

- Text document: ordered set of words
- Document collection
- „Bag of words“: unordered set of words
- Log files: designed for machine readability

→ Lecture on May 28th + 31st – TextVis

Dataset Availability

- Standard: **static files**

- Challenge today: **dynamic streams**

Tableau Example

- Let‘s create a map as example for spatial semantics
- Drag and drop the Country dimension into the center of a new sheet

Tableau Example

Tableau Example

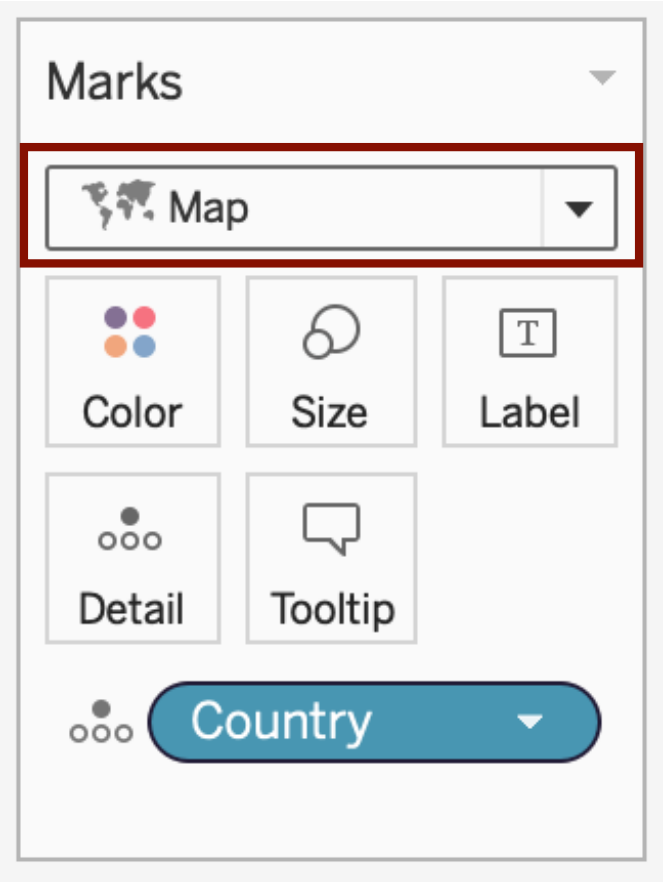


Tableau Example

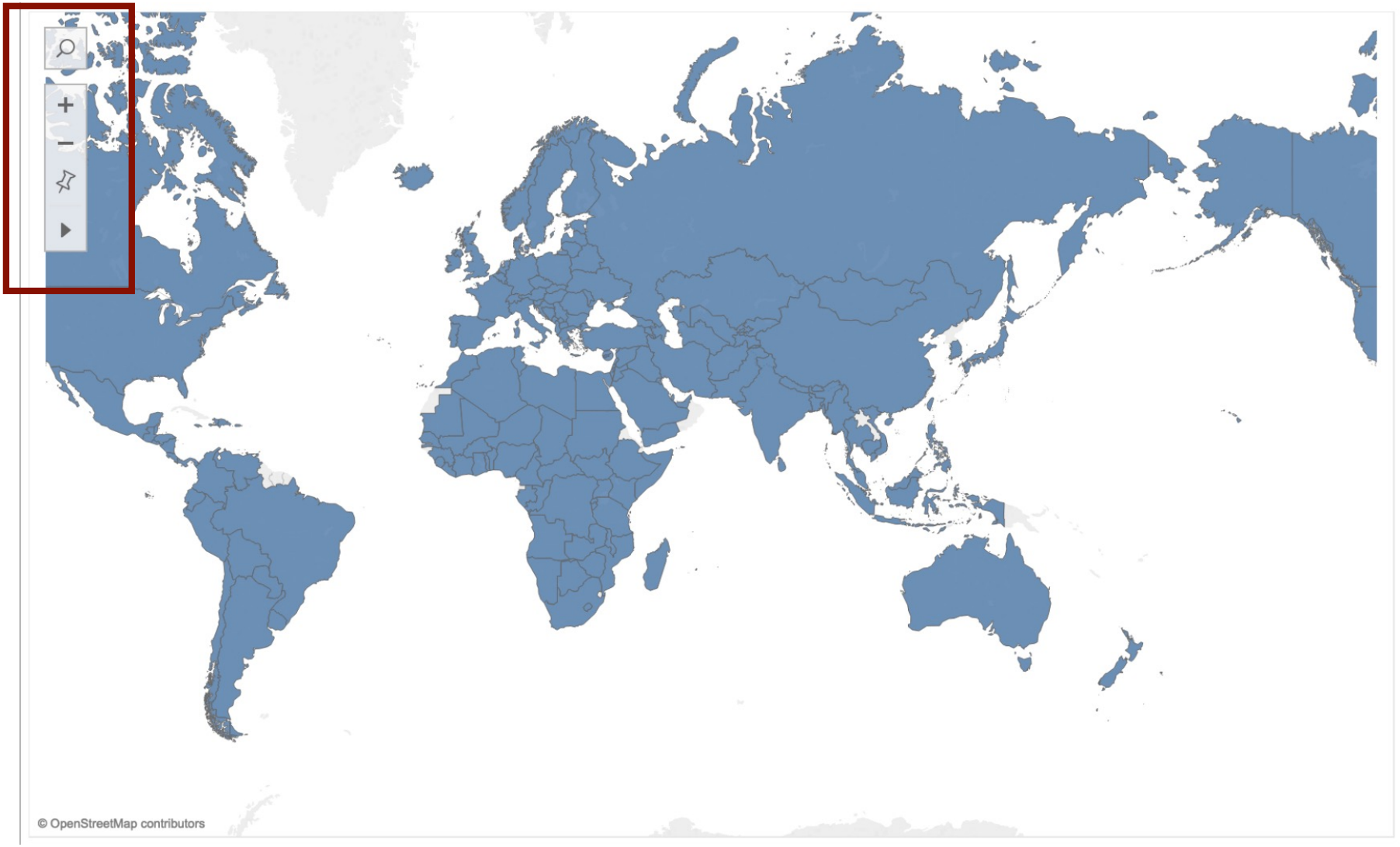


Tableau Example

- We can use this map now to explore different aspects of the dataset
- Let us drag the **Family Income Gini Coeff** field onto *Color*

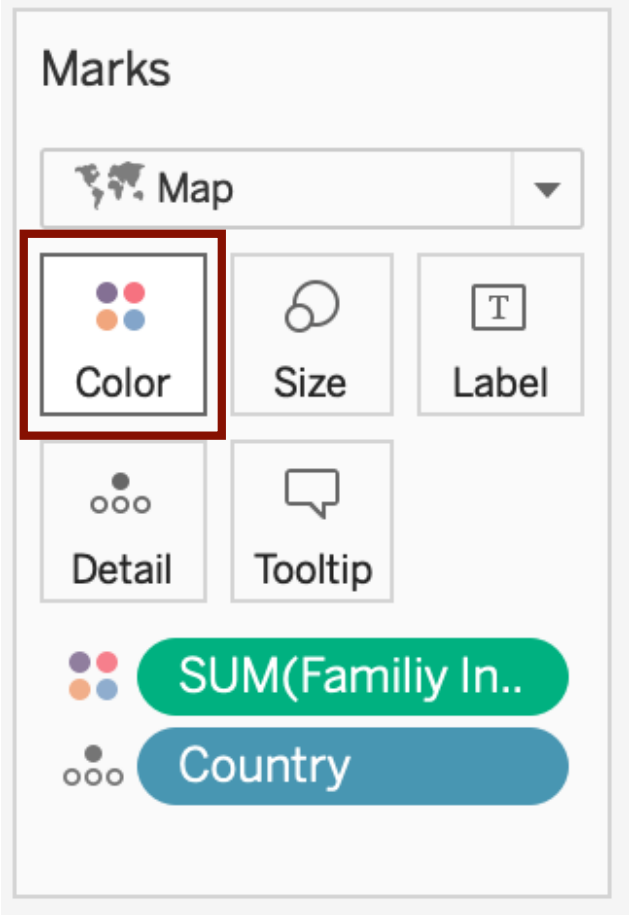
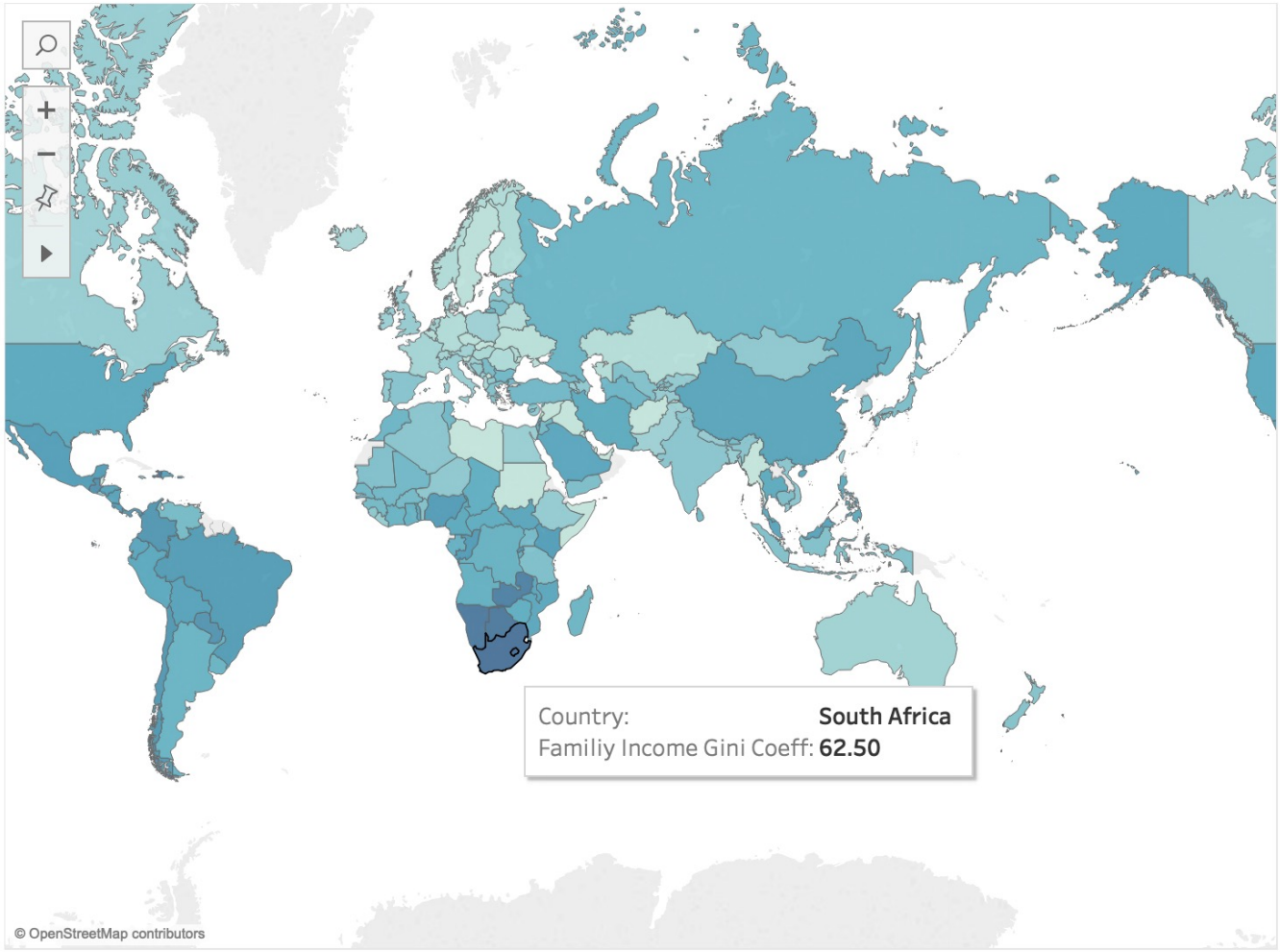


Tableau Example

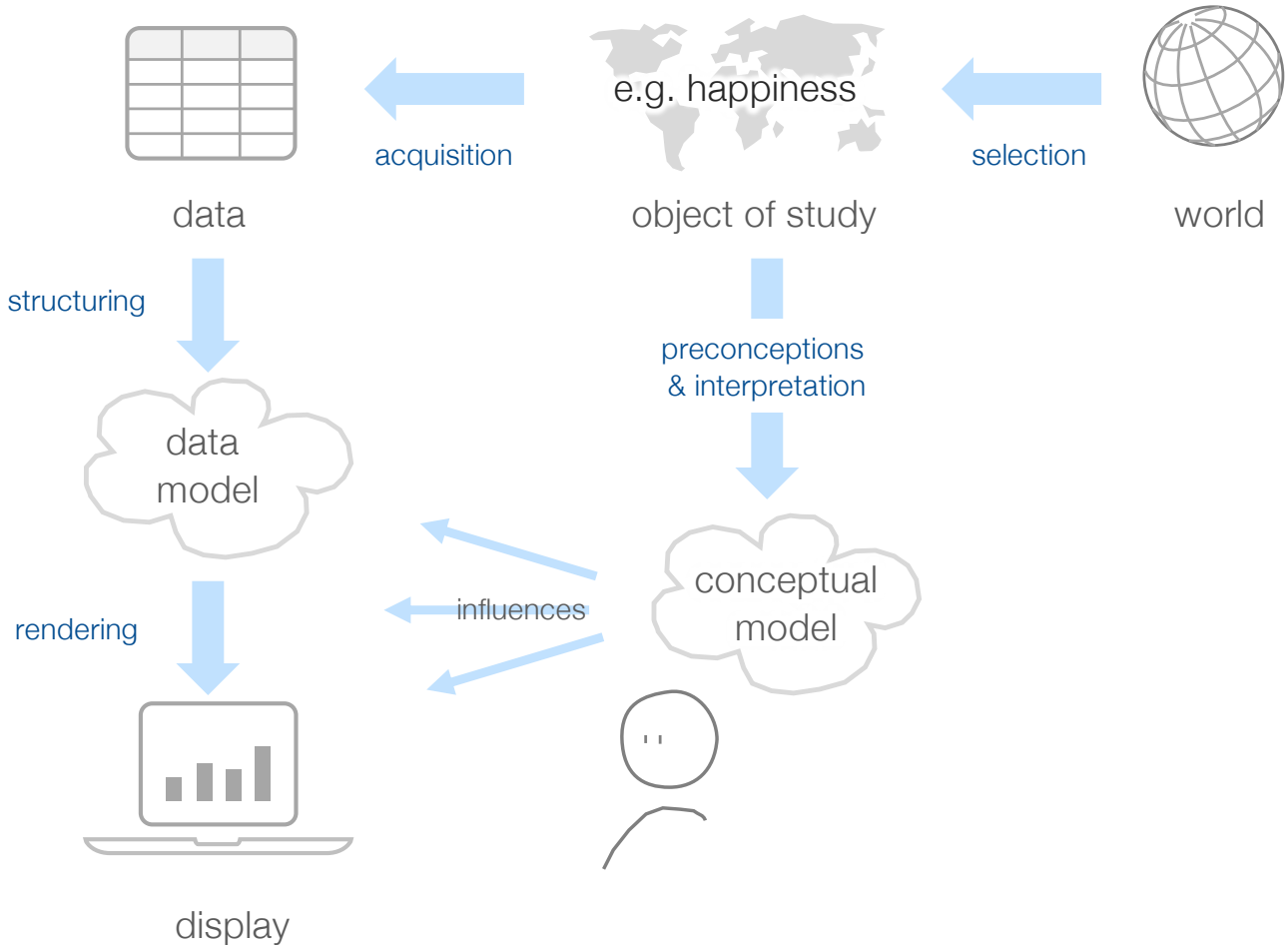


Attribute + Dataset Semantics

Attribute + Dataset Semantics

- Data vs. Concept
- Key vs. Value
  - High-dimensional, multi-dimensional, multi-variate
- Spatial vs. Non-spatial
- Temporal vs. Non-temporal
- Continuous vs. Discrete
- Topology vs. Geometry

Data vs. Conceptual Models



Data vs. Conceptual Models

- **Data Model**: Low-level description of the data
  - Set with operations, e.g., floats with +, -, /, *

- **Conceptual Model**: Mental construction
  - Includes semantics, supports reasoning

Physical Type	**Conceptual**
1D floats	temperature
3D vector of floats	space

Key vs. Value

Key = Index used to look up **value** attributes

- In "flat" table:
  - Implicit (row)
  - Explicit: contained within table as attribute
    - There must not be duplicates!
    - Keys may be categorical or ordinal (quantitative unsuitable)
- In multidimensional table:
  - Multiple keys required for item lookup
- In fields:
  - Spatial position acts as key
  - Characterized in terms of number of keys vs. values:
    - multi-variate structure depends on number of values
    - multi-dimensional structure depends on number of keys

High-dimensional vs. multi-dimensional vs. multi-variate

- Multi-dimensional
  Two to tens of dimensions
- High-dimensional:
  No constraint, can be thousands of dimensions and possibly more dimensions than samples
- Multi-variate:
  Multiple values per sample

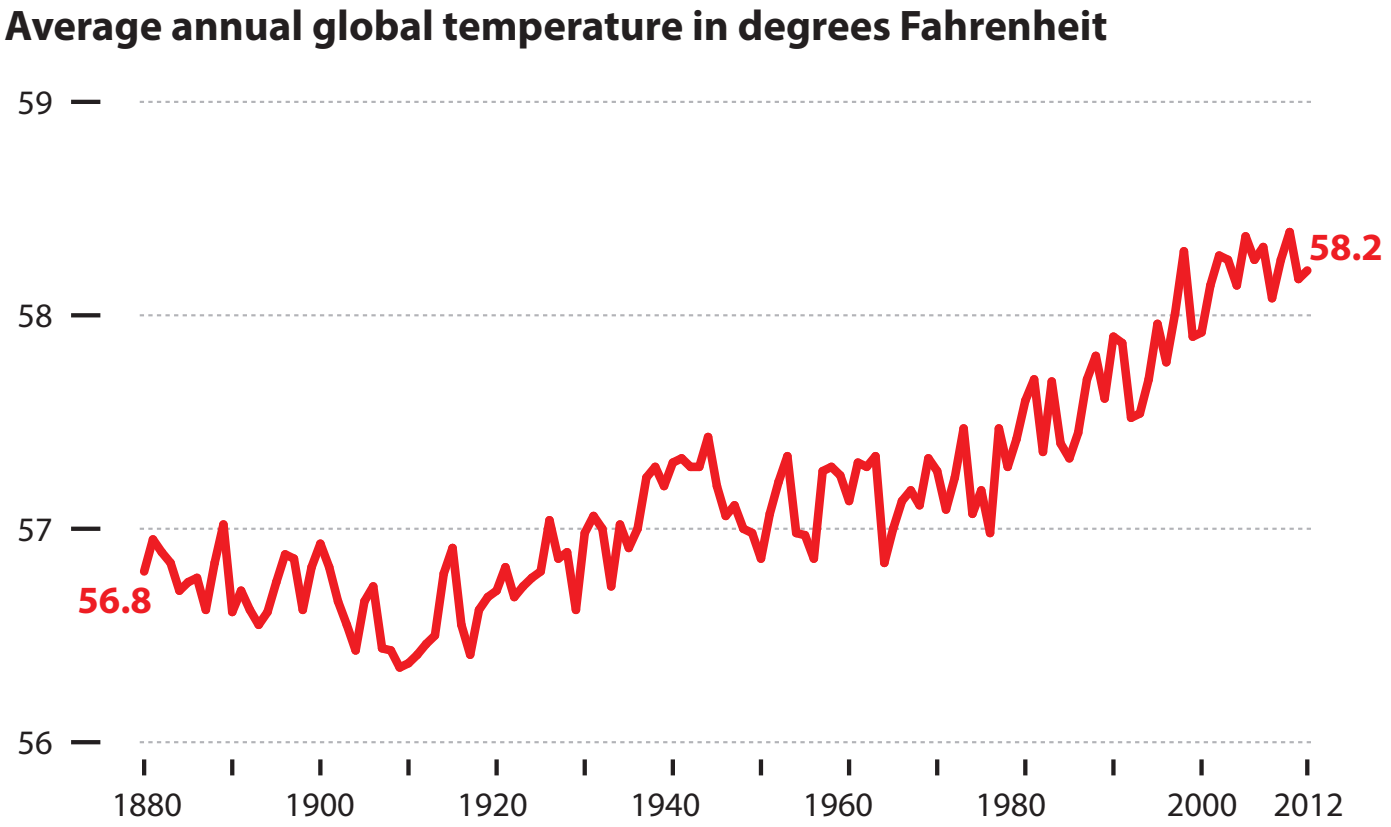
Spatial vs. Non-spatial / Abstract

- Implications on visual encoding
- Spatial
  - Geographic information
  - Physical simulation
  - Medical data (MRI, CT scan etc.)
  - Strong constraints on visual layout
- Non-spatial / abstract
  - Network data
  - Financial transactions
  - Up to the visualization expert to choose a visual layout

Temporal / Time-varying vs. Non-temporal / Static

- Time has a strong meaning to us as humans
- Special consideration for visual encoding
- Time is multi-scale and has a hierarchy
  (… minutes - hours - days - weeks …)
- Time periods/cycles very important
- Can have either value or key semantics
- Time series dataset
  - Special case of table where time is key
  - Time-value pairs often at uniform temporal intervals
  - Typical tasks: Find trends, correlations and variations at multiple time scales

Time Series: Example



Alberto Cairo, How charts lie, 2019

Continuous vs. discrete

- Data is almost always discrete –
  we need to store it in discrete memory cells
- It's really how we think about the data
- Categorical is always discrete
- Quantitative is continuous
- Care must be taken when making
  discrete measurements continuous

Continuous vs. discrete: Example

- From data model...
  - 32.5, 54.0, -17.3, … (floats)
- Using conceptual model...
  - Temperature
- To data type
  - Continuous to 4 significant digits (Q)
  - Hot, warm, cold (O)
  - Burned vs. Not burned (N)

Topology vs. Geometry

- **Topology**
  specifies the structure (**connectivity**) of the data

- **Geometry**
  specifies the **position** of the data

Topology vs. Geometry

- In **topology**, qualitative questions about geometrical structures are the main concern
  - Does it have any holes in it?
  - Is it all connected together?
  - Can it be separated into parts?

- Underground maps do not show how far one station is from the other, but rather how they are connected (=topological map)

Derived Data

Derived Attributes

- The norm, not the exception
- Necessary for some of the tasks
- Simple transformations
- Statistical summaries of (lots of) data

→ Lecture on June 11th – Reduce Items & Attributes

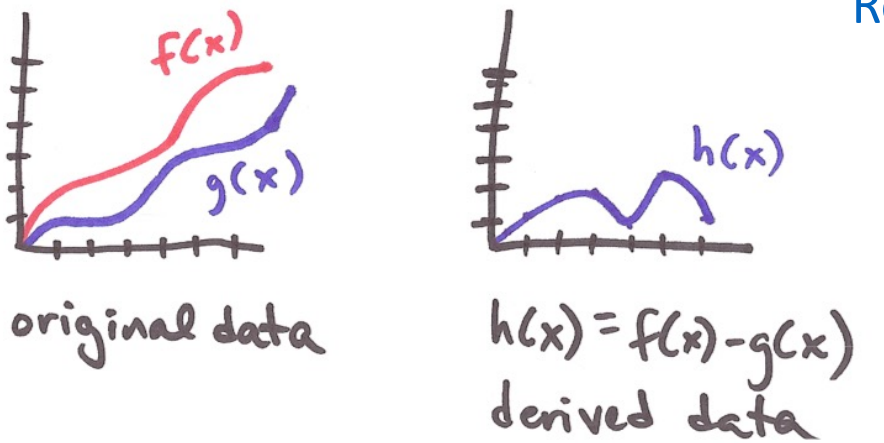
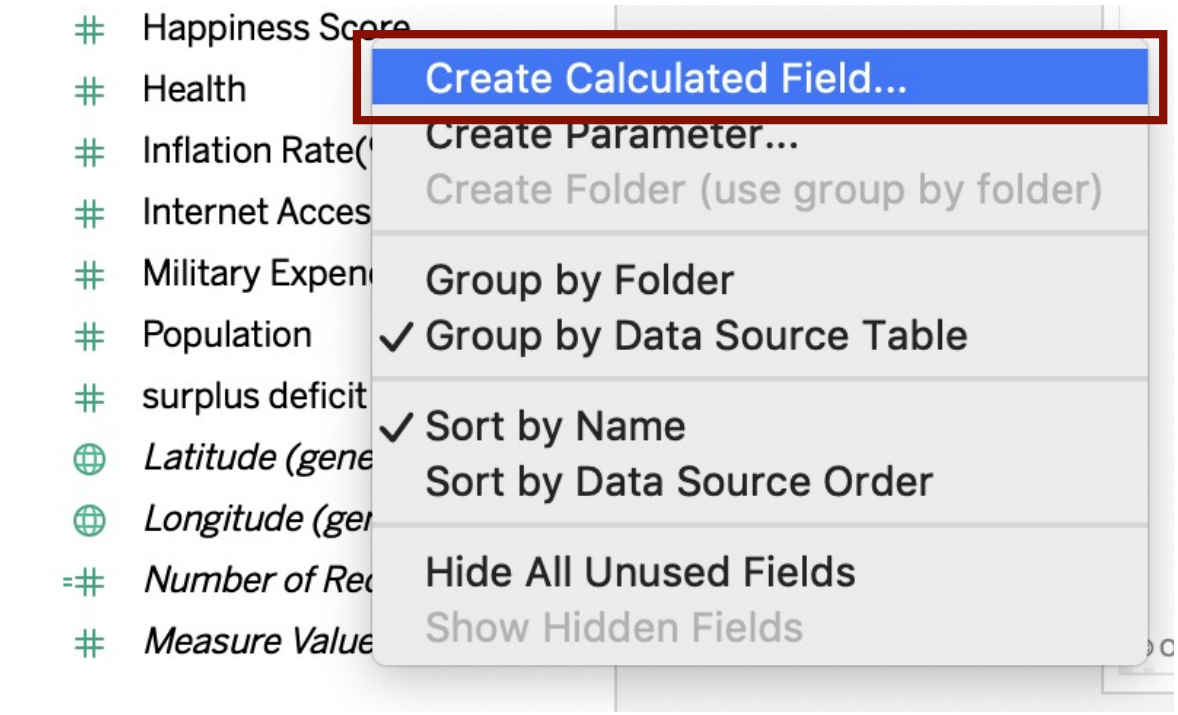


Tableau Example

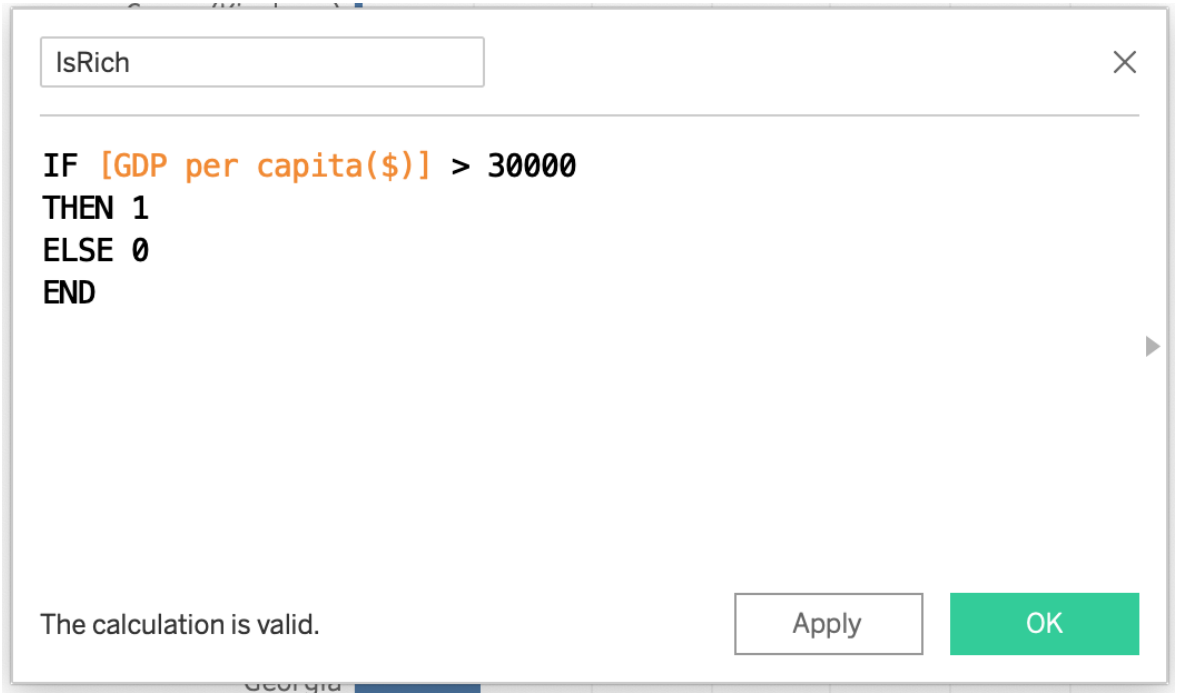
An easy way to create your own derived attributes are Calculated Fields in Tableau

Tableau Example



Right-Click somewhere in the white space below the *Measures* and create a Calculated Field

Tableau Example



Lets drag that new Measure **IsRich** onto *Color* in the Map Sheet

Tableau Example

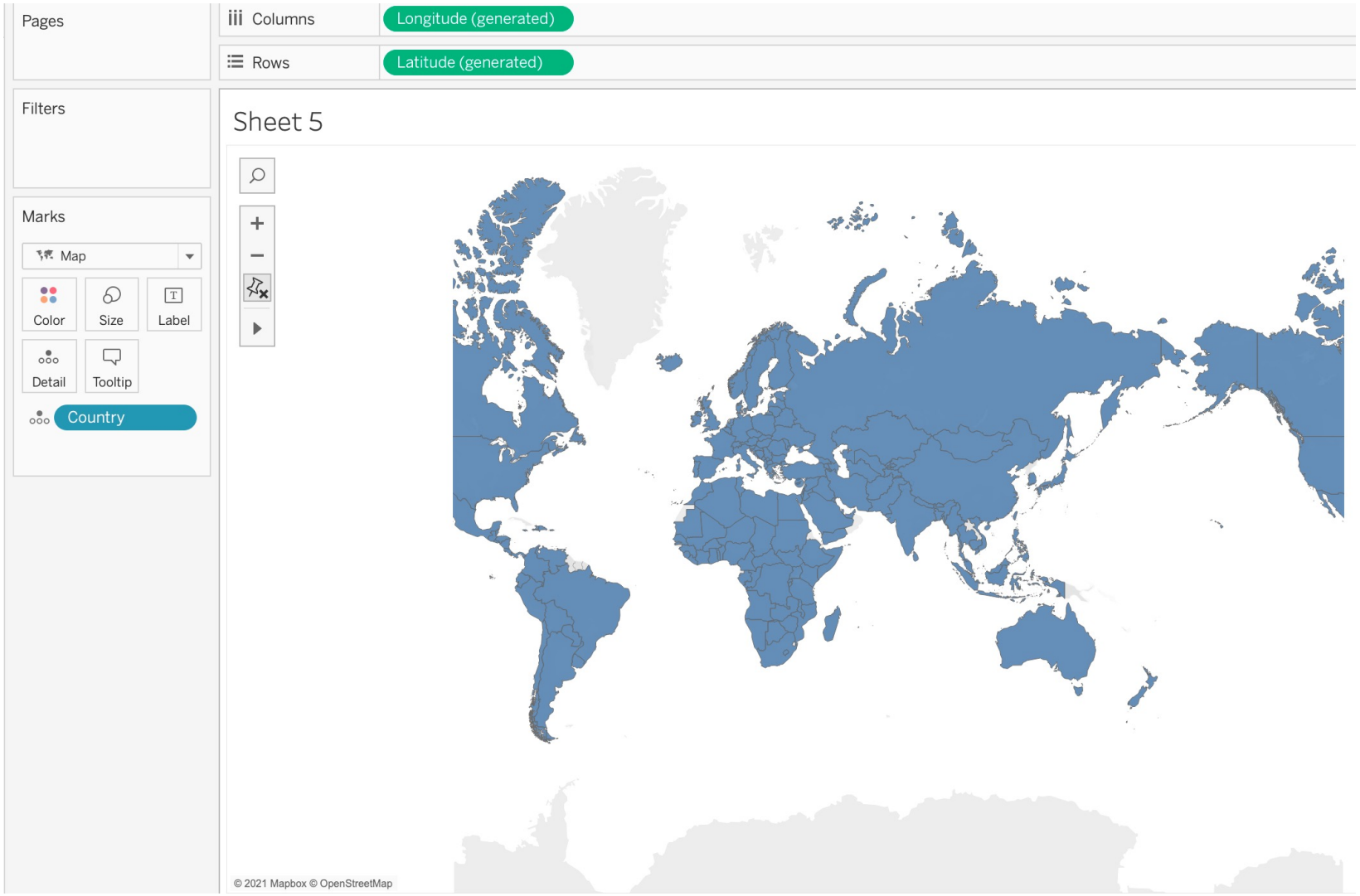
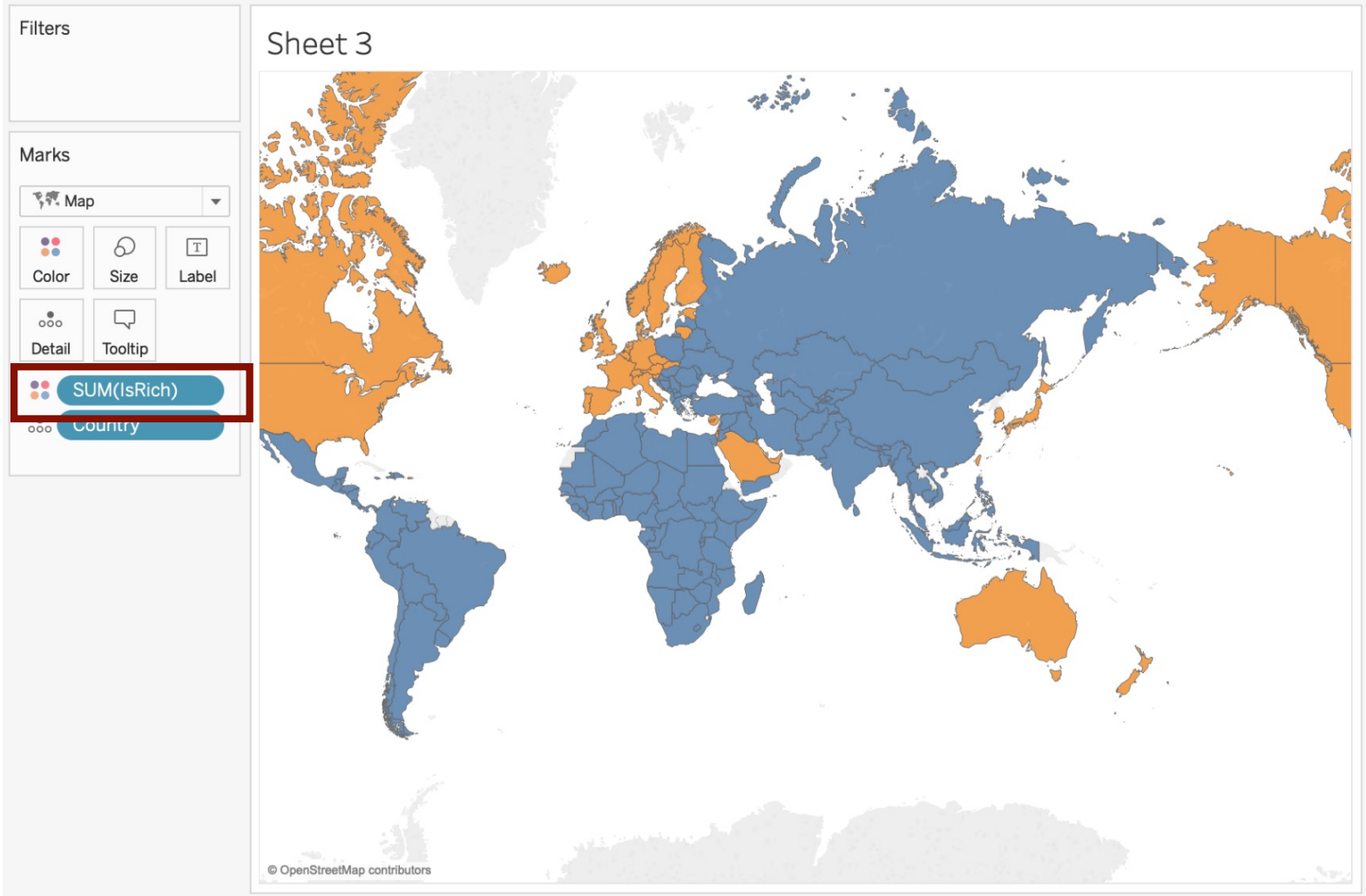


Tableau Example

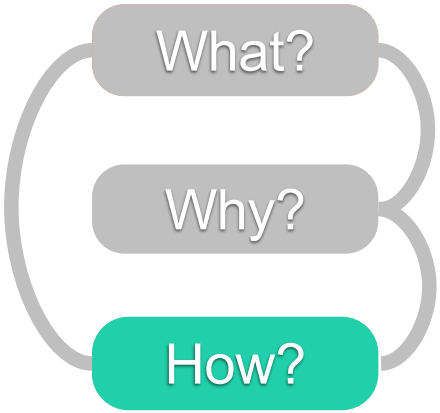


Summary

Data abstraction helps in thinking which visualization to choose / how tasks can be solved!

- Semantics
  vs. Data(set) types
- Attribute Types
- Data Types
- Dataset Types
- Attribute and Dataset Semantics

Outlook



What data is shown?

Why is the task being performed?

How is the visualization constructed?