Towards Data Visualisation based on Conceptual Modelling

P.J. McBrien and A. Poulovassilis

ER 2018

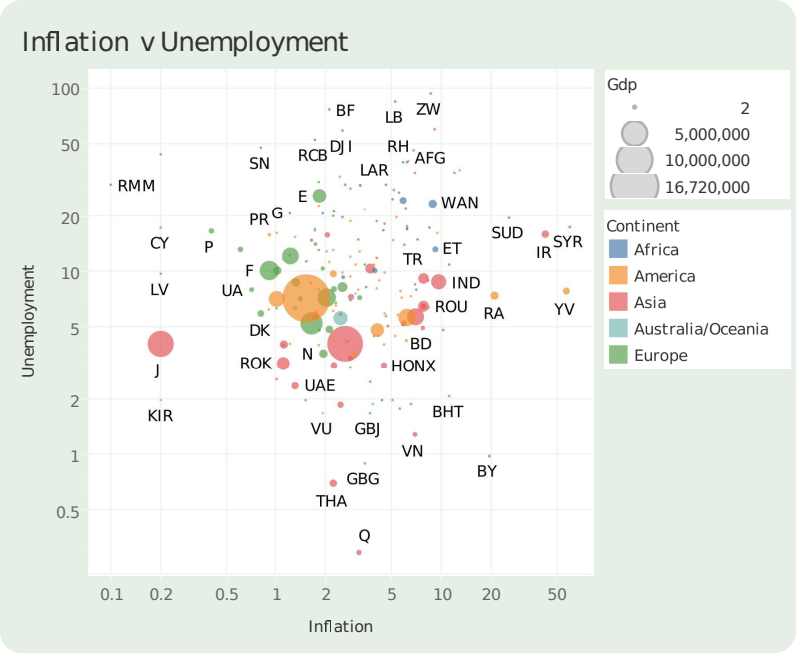
Motivating Example

continent country economy							
name	code	population	unemployment	inflation	gdp	continent	percentage
Argentina	RA	42669500	7.50	20.80	484600	America	100
Australia	AUS	23135281	5.70	2.40	1488000	Australia/Oceania	100
Belize	BZ	312971	15.50	1.30	1637	America	100
Bolivia	BOL	10027262	7.40	6.50	30790	America	100
Brazil	BR	202768562	5.70	6.20	2190000	America	100
Canada	CDN	35158304	7.10	1	1825000	America	100
Chile	RCH	16341929	6	1.70	281700	America	100
China	CN	1360720000	4.10	2.60	9330000	Asia	100
Costa Rica	CR	4773119	7.90	5.60	48510	America	100
India	IND	1210854977	8.80	9.60	1670000	Asia	100
Panama	PA	3405813	4.50	4.10	40620	America	100
Paraguay	PY	6672631	6.60	2.30	30560	America	100
Peru	PE	30135875	3.60	2.90	210300	America	100
Russia	R	143666931	5.80	6.80	2113000	Europe	25
Russia	R	143666931	5.80	6.80	2113000	Asia	75
United Kingdom	GB	64105654	7.20	2	2490000	Europe	100
United States	USA	318857056	7.30	1.50	16720000	America	100
Uruguay	ROU	3286314	6.50	8.30	57110	America	100

Current approaches require users to build visualisations from tables of data.

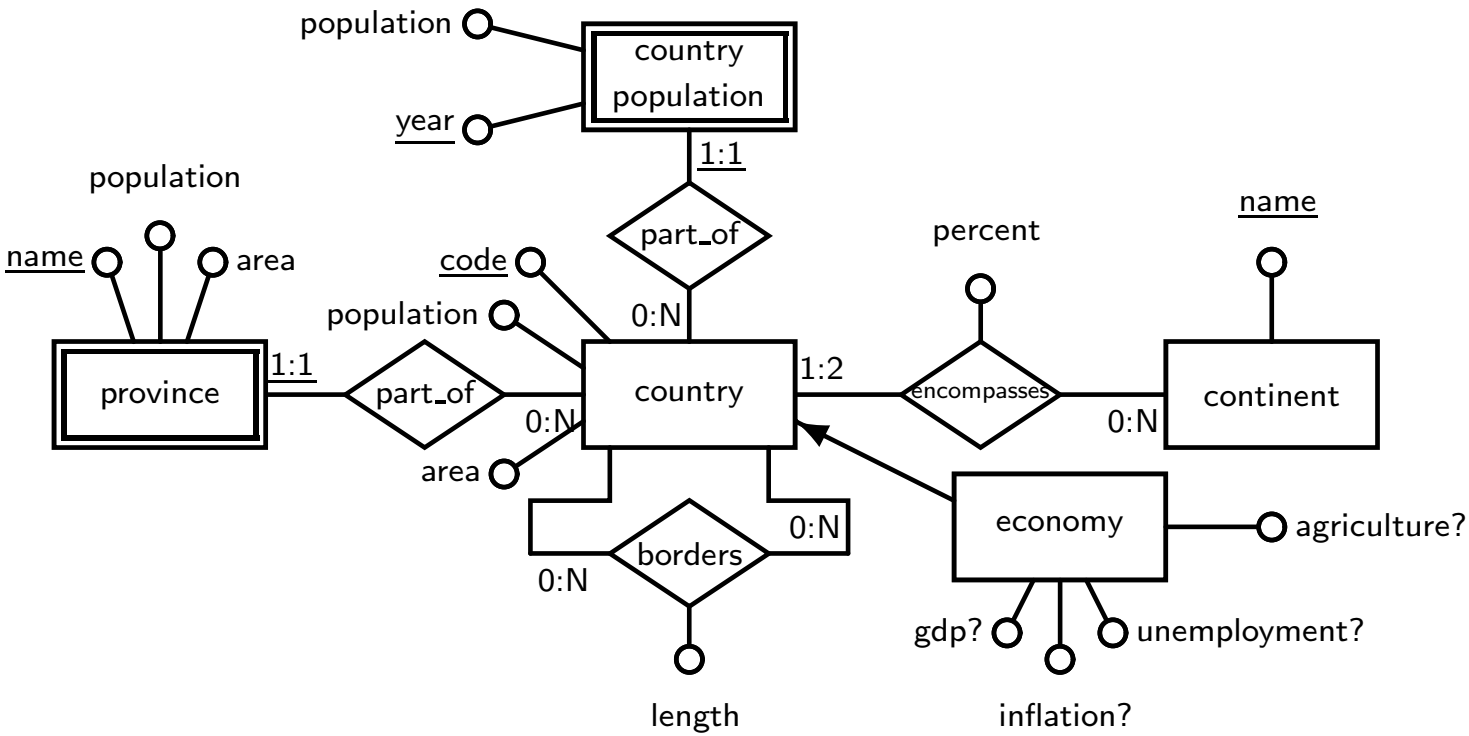
Motivating Example

continent country economy							
name	code	population	unemployment	inflation	gdp	continent	percentage
Argentina	RA	42669500	7.50	20.80	484600	America	100
Australia	AUS	23135281	5.70	2.40	1488000	Australia/Oceania	100
Belize	BZ	312971	15.50	1.30			
Bolivia	BOL	10027262	7.40	6.50			
Brazil	BR	202768562	5.70	6.20			
Canada	CDN	35158304	7.10	1			
Chile	RCH	16341929	6	1.70			
China	CN	1360720000	4.10	2.60			
Costa Rica	CR	4773119	7.90	5.60			
India	IND	1210854977	8.80	9.60			
Panama	PA	3405813	4.50	4.10			
Paraguay	PY	6672631	6.60	2.30			
Peru	PE	30135875	3.60	2.90			
Russia	R	143666931	5.80	6.80			
Russia	R	143666931	5.80	6.80			
United Kingdom	GB	64105654	7.20	2			
United States	USA	318857056	7.30	1.50			
Uruguay	ROU	3286314	6.50	8.30			



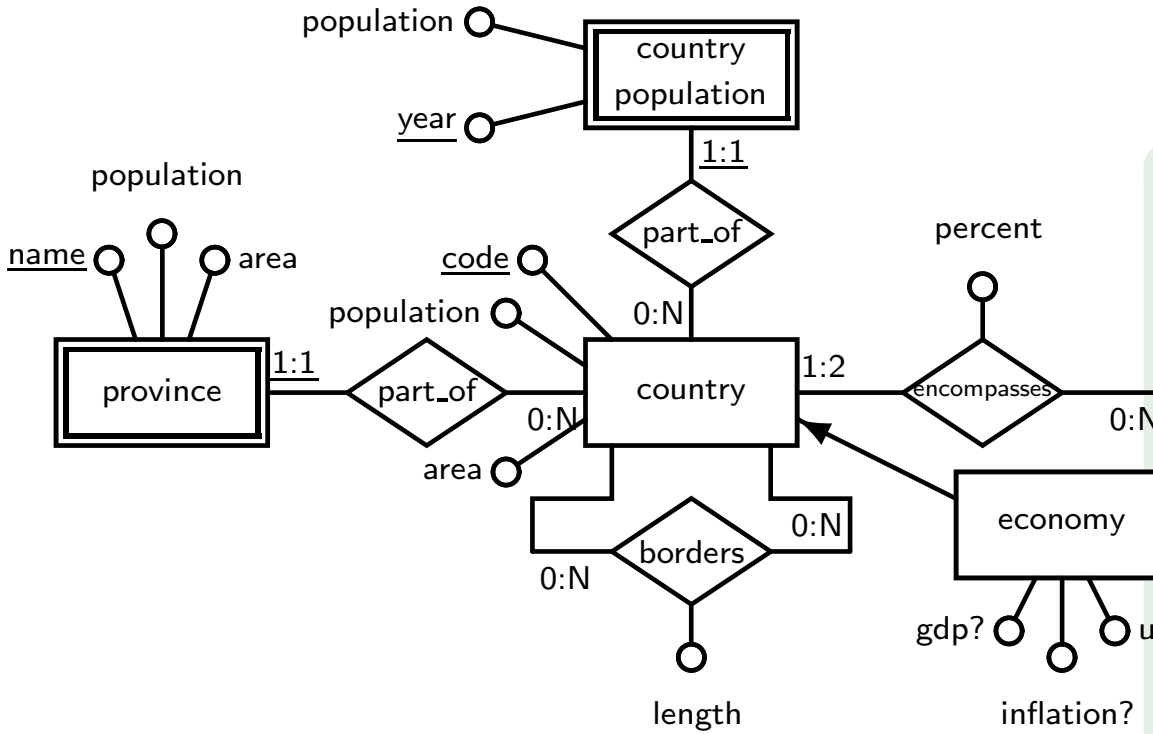
Considerable effort is needed to understand the data and decide on which visualisation.

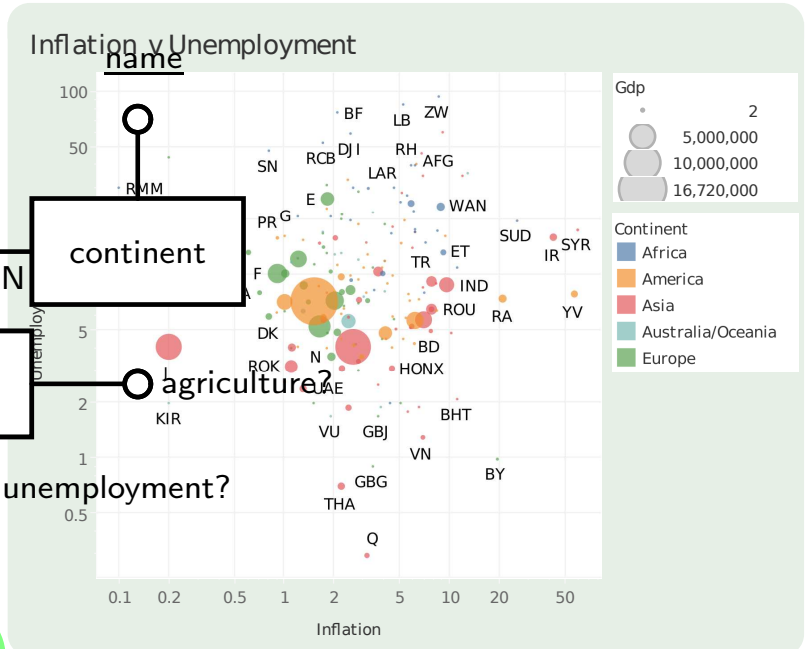
Motivating Example



Our approach uses the conceptual model of the data to semi-automatically aid the user's choice of visualisations.

Motivating Example





Our approach uses the conceptual model of the data to semi-automatically aid the user's choice of visualisations.

Dimensions of a Visualisation

The elements of a visualisation are often classified in the Visualisation literature as:

- **marks** *e.g.* points, lines, areas

Dimensions of a Visualisation

The elements of a visualisation are often classified in the Visualisation literature as:

- **marks** *e.g.* points, lines, areas
- **channels** *e.g.* colour, length, shape, coordinate, texture

Dimensions of a Visualisation

The elements of a visualisation are often classified in the Visualisation literature as:

- **marks** *e.g.* points, lines, areas
- **channels** *e.g.* colour, length, shape, coordinate, texture

In our approach, each attribute of an entity and each participation of an entity in a relationship is associated with a dimension of the visualisation. Dimensions are of two types:

- **discrete dimensions** which map to a mark or a channel

Discrete Dimension Example

continent.name={'Europe','Asia','Australia/Oceania','Africa','America'}

Dimensions of a Visualisation

The elements of a visualisation are often classified in the Visualisation literature as:

- **marks** *e.g.* points, lines, areas
- **channels** *e.g.* colour, length, shape, coordinate, texture

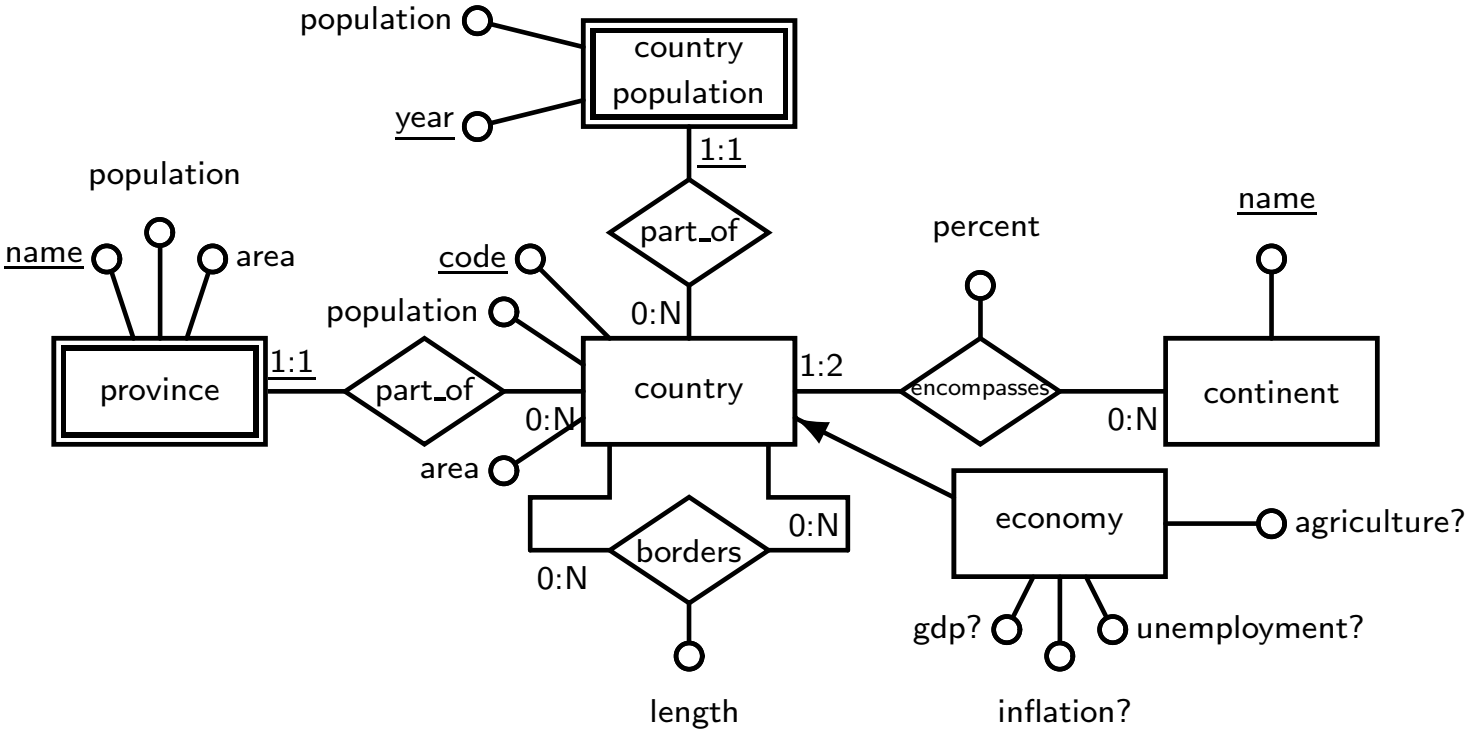
In our approach, each attribute of an entity and each participation of an entity in a relationship is associated with a dimension of the visualisation. Dimensions are of two types:

- **discrete dimensions** which map to a mark or a channel
- **scalar dimensions** which map to a channel

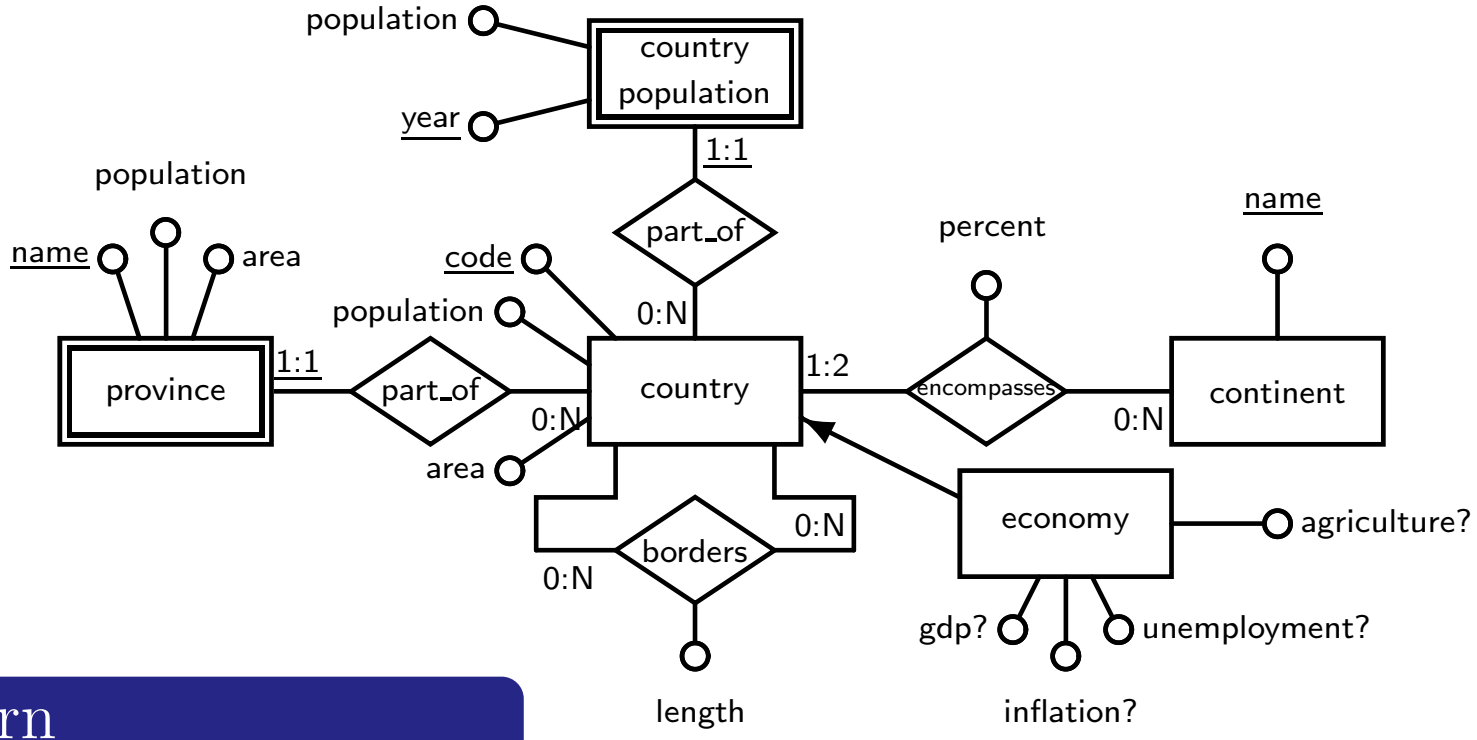
Scalar Dimension Example

borders.length={0.34,1.20,3.2,4.40,6.30.9.00,9.60,. . . ,5150,6846,8893}

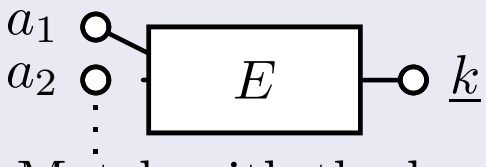
Visualisation Schema Pattern 1: Basic Entity



Visualisation Schema Pattern 1: Basic Entity

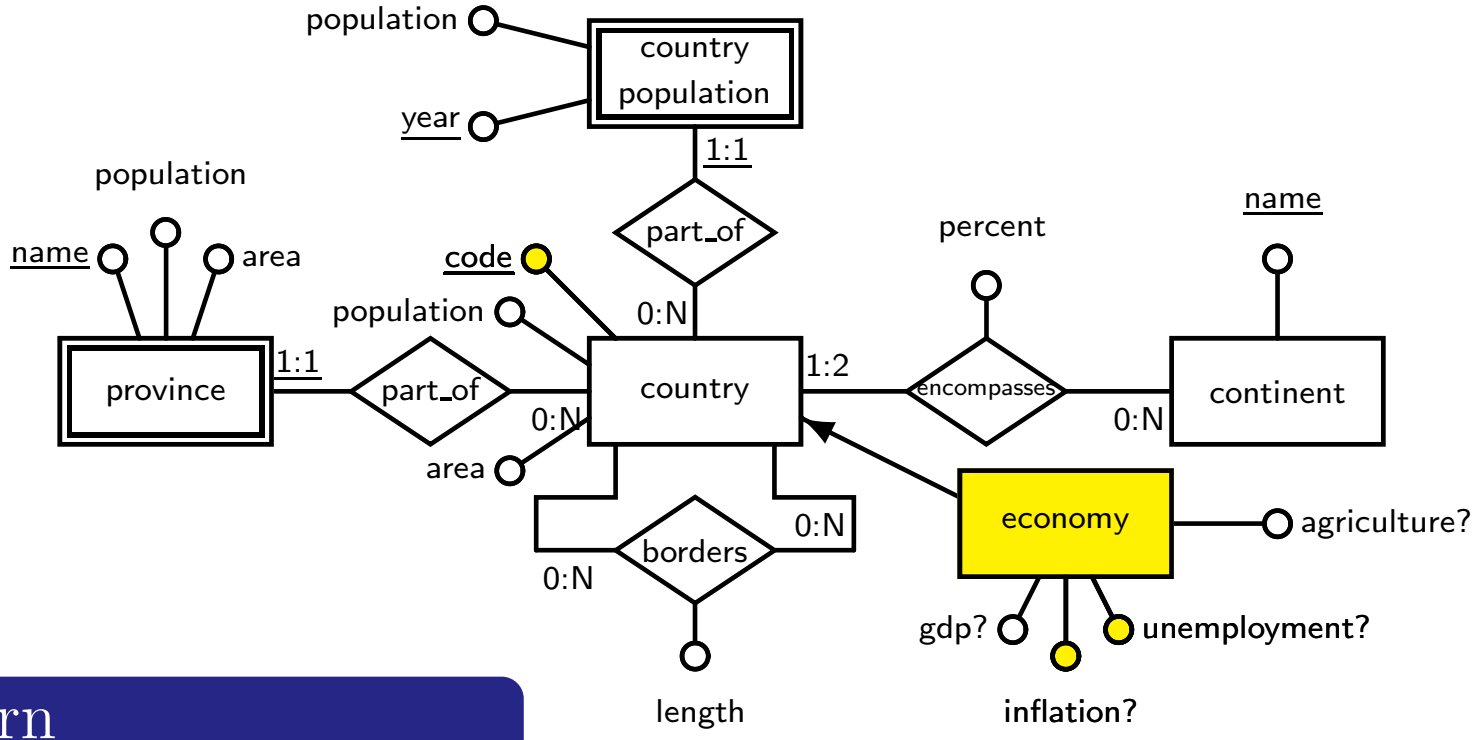


Pattern

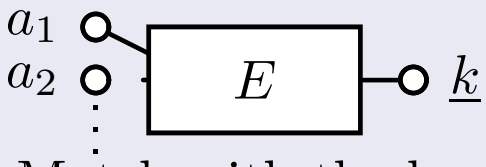


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 1: Basic Entity

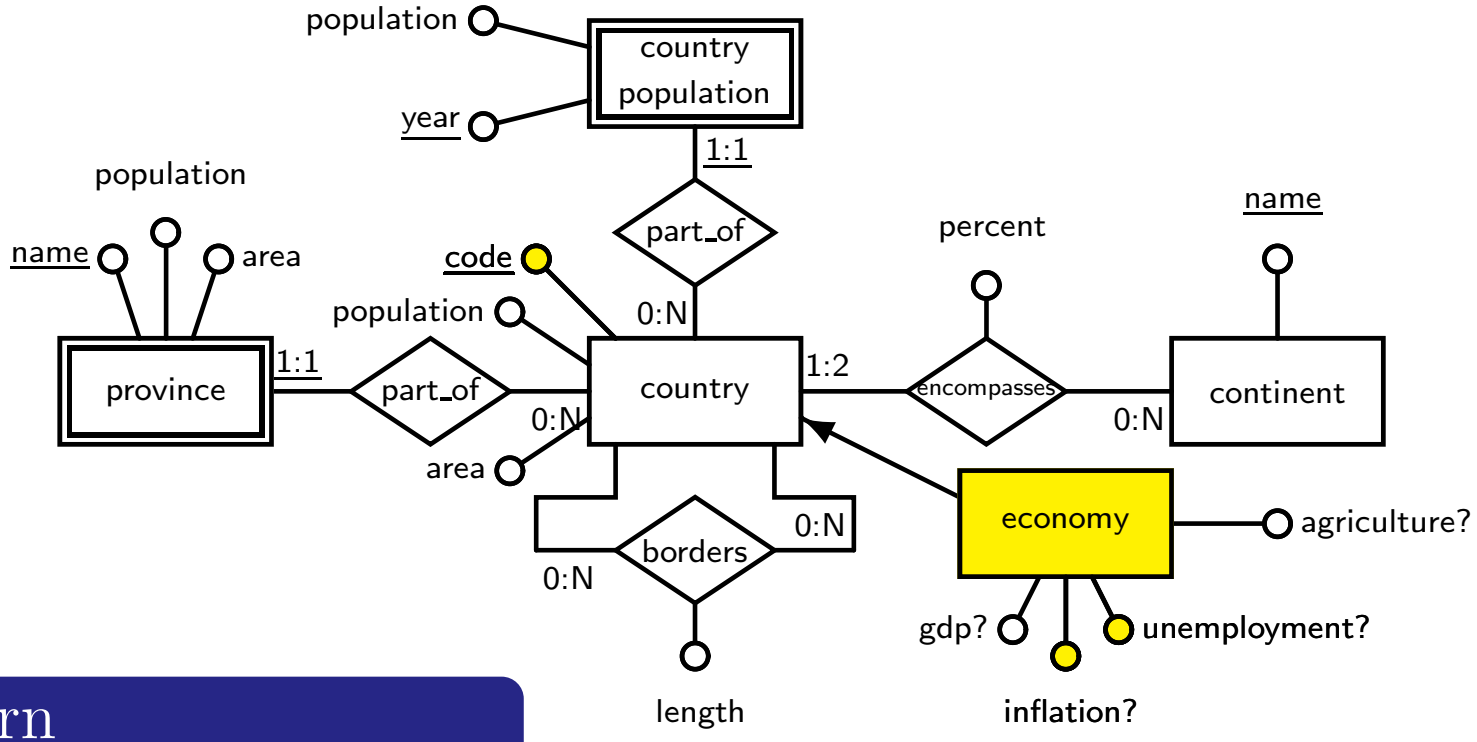


Pattern

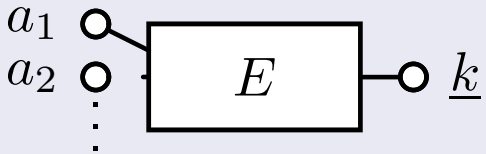


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 1: Basic Entity

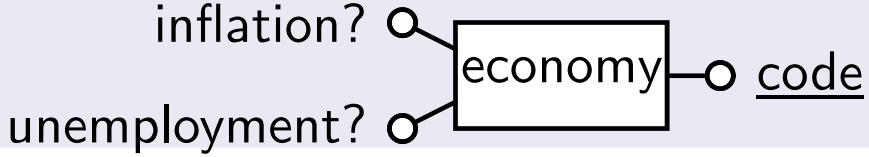


Pattern



Match with the key of an entity, plus one or more attributes

Instantiation

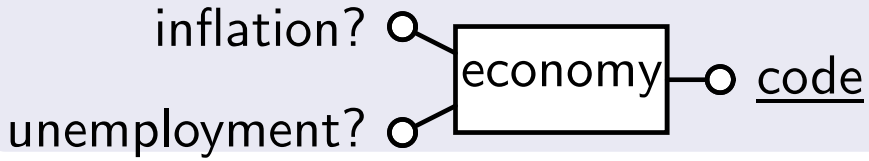


Visualisation Schema Pattern 1: Basic Entity

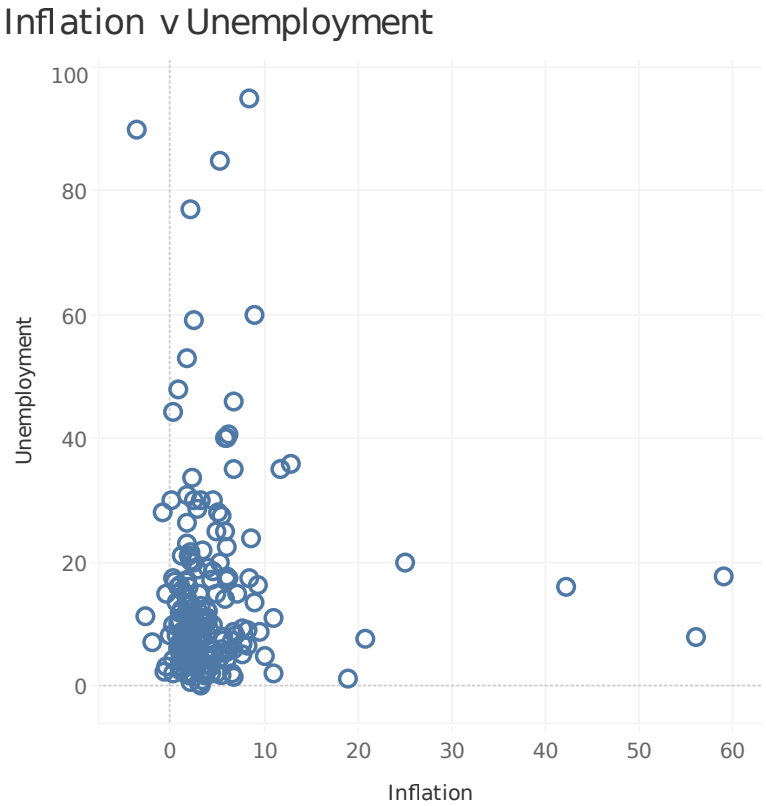
Visualisation Types

Basic Entity Visualisations			
Name	$\|k\|$	mandatory	optional
Bar Chart	1..100	a_1 scalar	-
Calendar	1..*	a_1 temporal scalar	a_2 colour
Scatter Diagrams	1..*	a_1, a_2 scalar	a_3 colour
Bubble Charts	1..*	a_1, a_2, a_3 scalar	a_4 colour
Choropleth Maps	1..*	k geographical, a_1 colour	-
Word Clouds	1..*	k lexical, a_1 scalar	a_2 colour

Instantiation



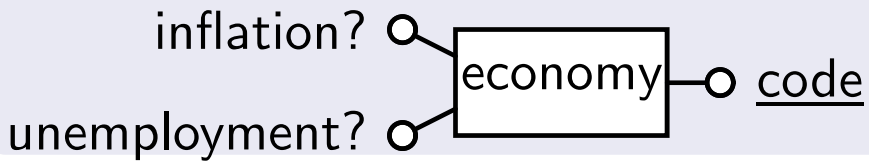
Visualisation Schema Pattern 1: Basic Entity



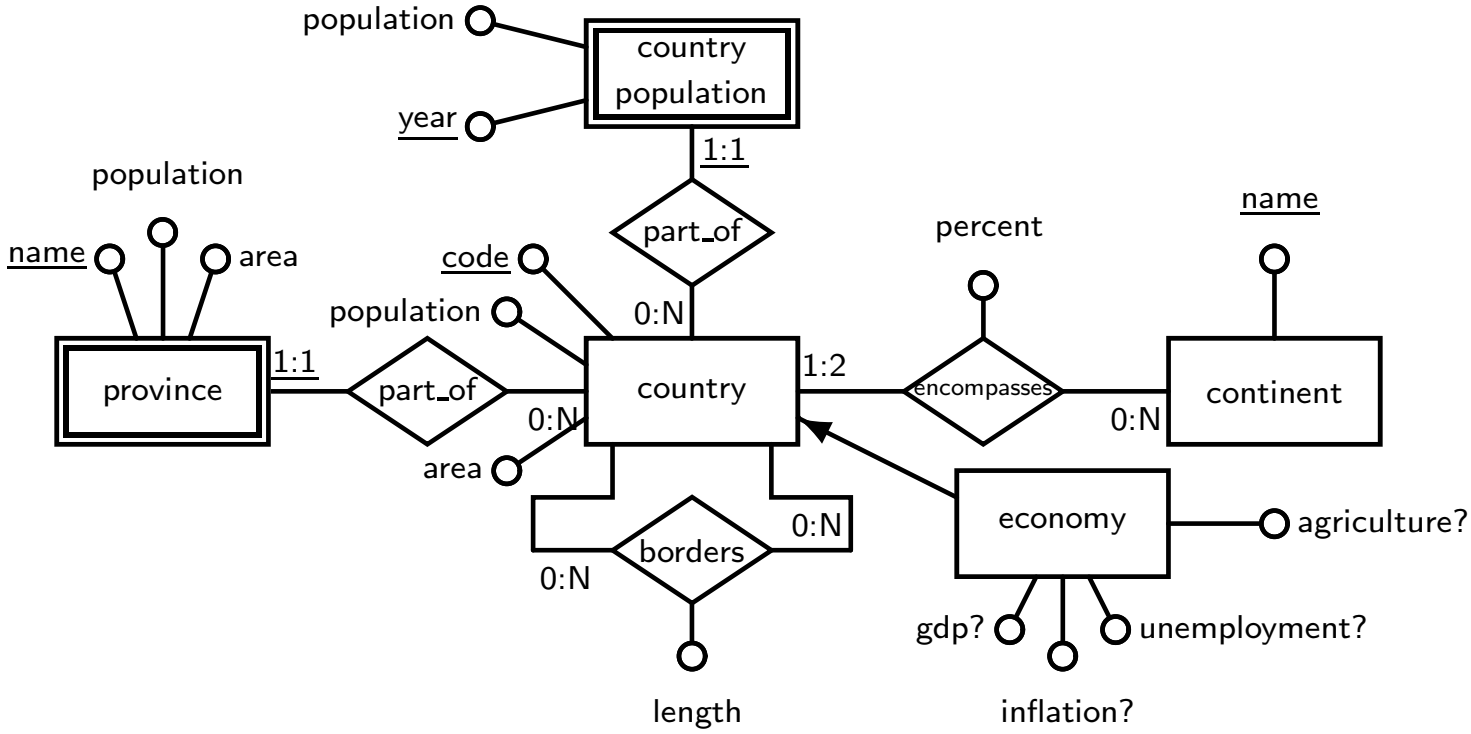
Visualisation Types

Basic Entity Visualisations			
Name	$\|k\|$	mandatory	optional
Bar Chart	1..100	a_1 scalar	-
Calendar	1..*	a_1 temporal scalar	a_2 colour
Scatter Diagrams	1..*	a_1, a_2 scalar	a_3 colour
Bubble Charts	1..*	a_1, a_2, a_3 scalar	a_4 colour
Choropleth Maps	1..*	k geographical, a_1 colour	-
Word Clouds	1..*	k lexical, a_1 scalar	a_2 colour

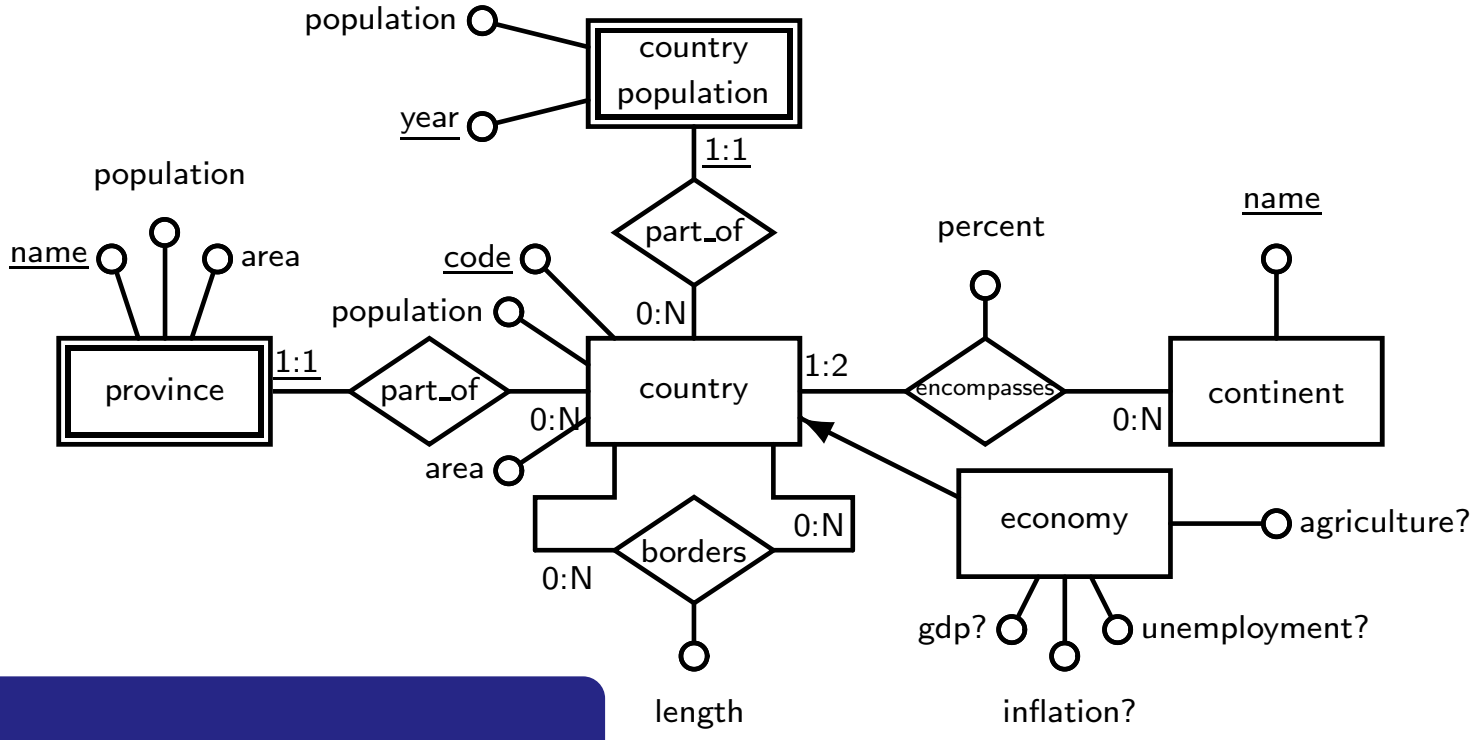
Instantiation



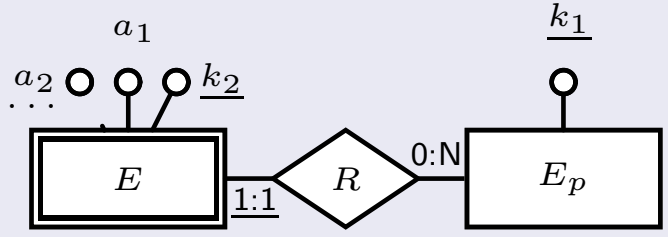
Visualisation Schema Pattern 2: Weak Entity



Visualisation Schema Pattern 2: Weak Entity

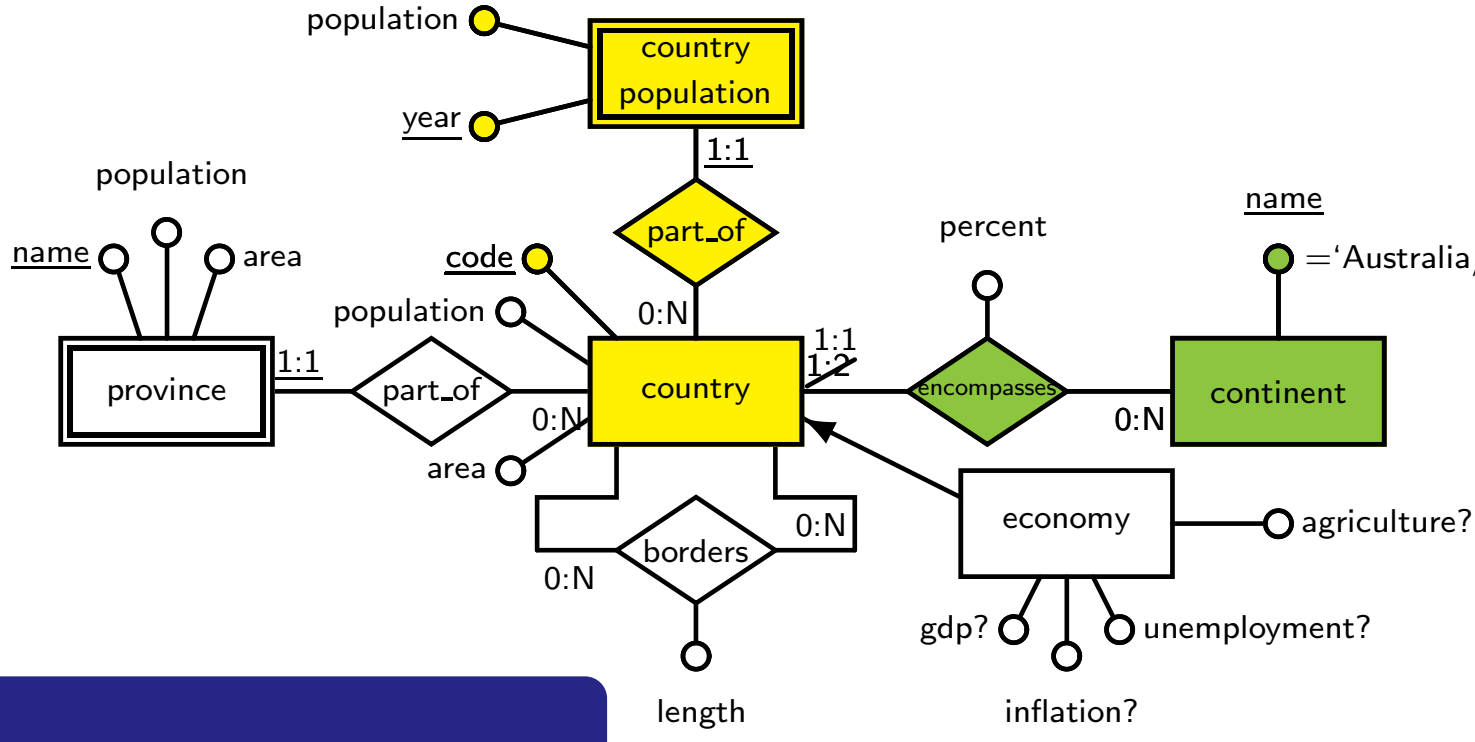


Pattern

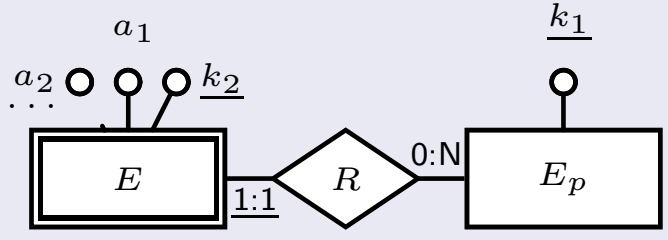


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 2: Weak Entity

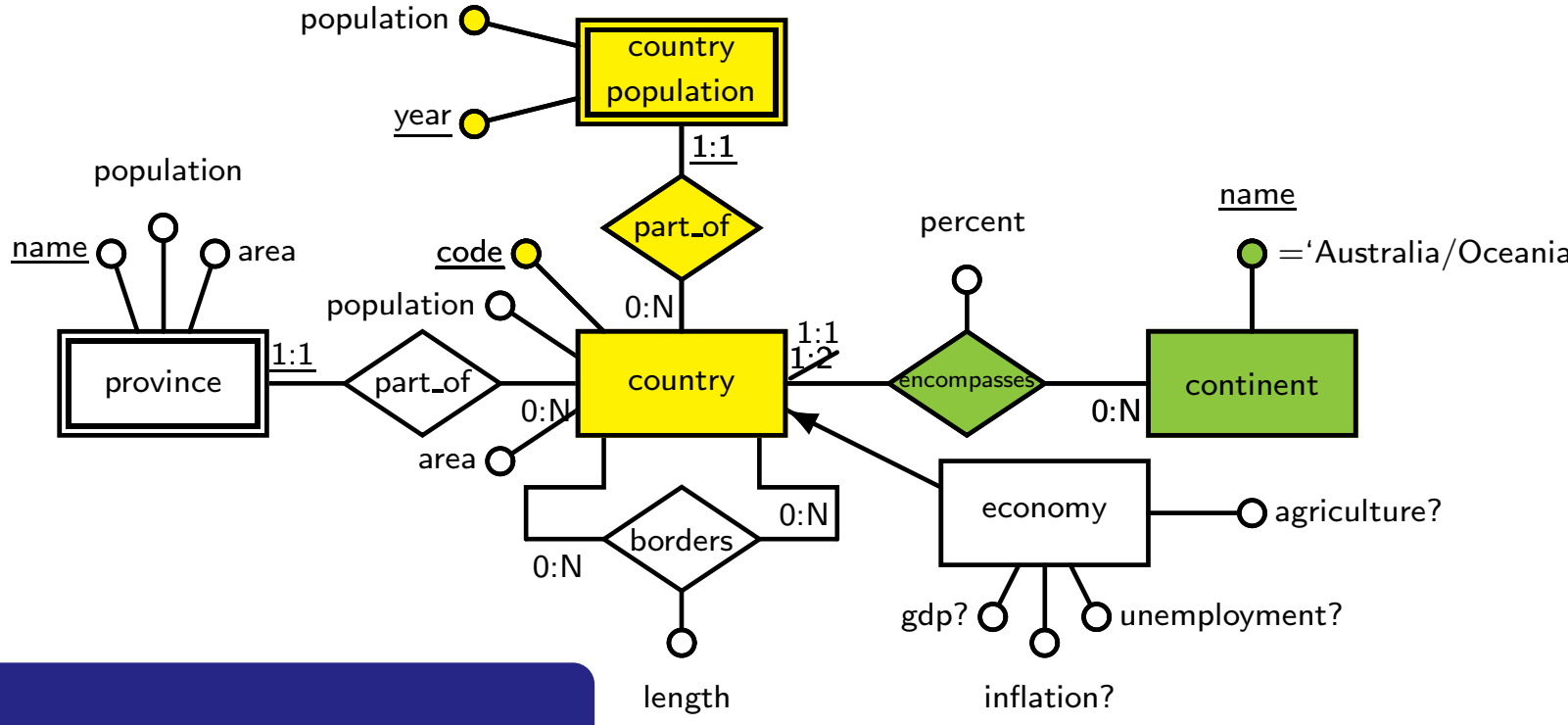


Pattern

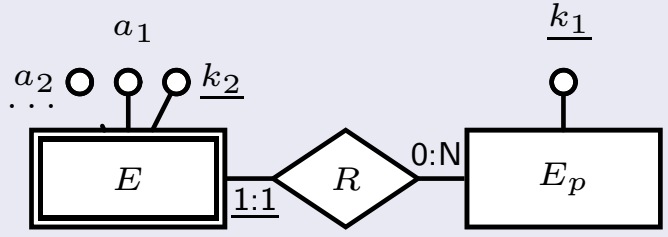


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 2: Weak Entity

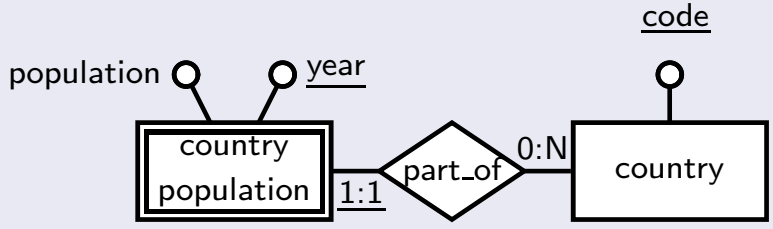


Pattern



Match with the key of an entity, plus one or more attributes

Instantiation

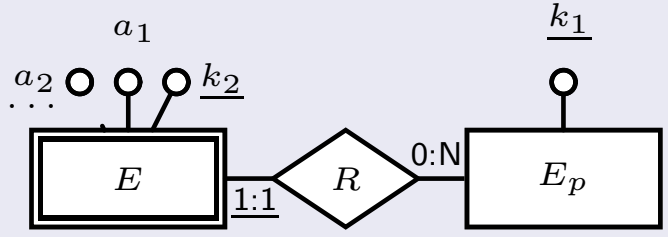


Visualisation Schema Pattern 2: Weak Entity

Visualisation Types

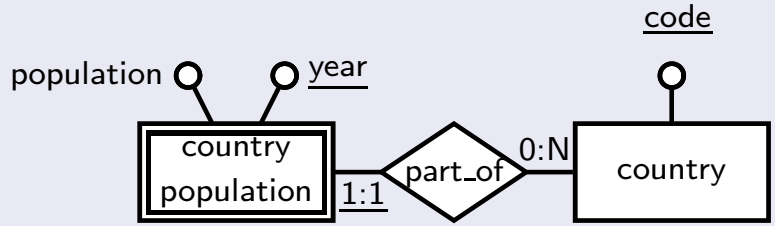
Weak Entity Visualisations					
Name	$\|k_1\|$	$\|k_2\|$	complete	mandatory	optiona
Line Chart	1..20	1..*	no	k_2, a_1 scalar	a_2 scal
Stacked Bar Chart	1..20	1..20	yes	a_1 scalar	-
Grouped Bar Chart	1..20	1..20	no	a_1 scalar	-
Spider Chart	3..10	1..20	yes	a_1 scalar	-

Pattern

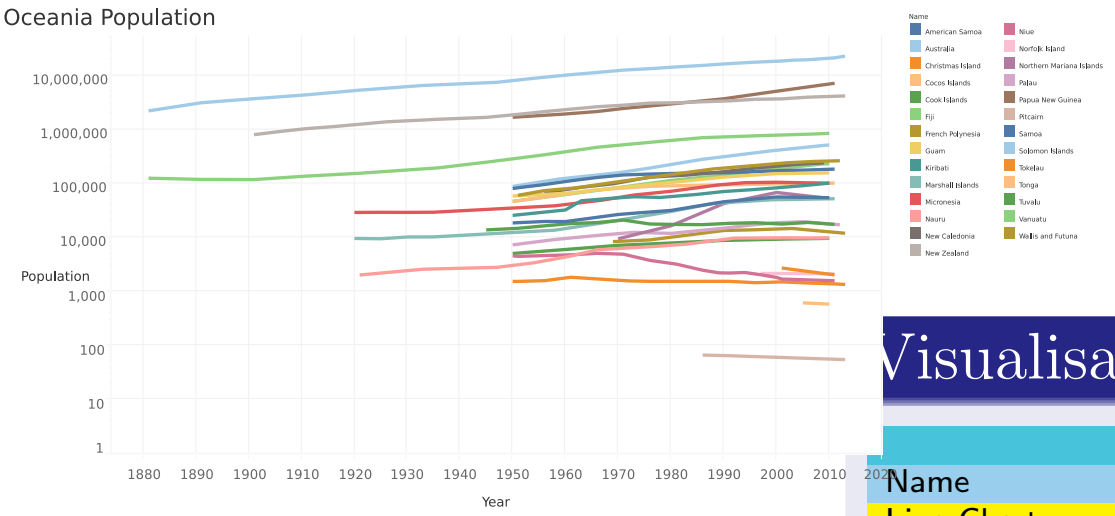


Match with the key of an entity, plus one or more attributes

Instantiation



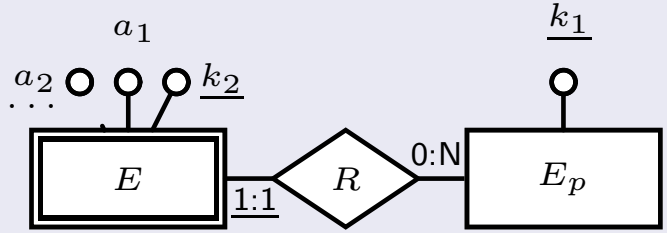
Visualisation Schema Pattern 2: Weak Entity



Visualisation Types

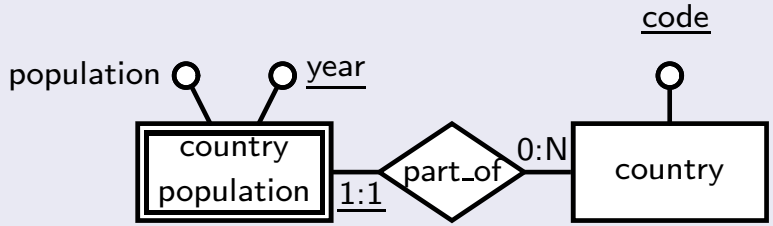
Weak Entity Visualisations					
Name	$\|k_1\|$	$\|k_2\|$	complete	mandatory	optiona
Line Chart	1..20	1..*	no	k_2, a_1 scalar	a_2 scal
Stacked Bar Chart	1..20	1..20	yes	a_1 scalar	-
Grouped Bar Chart	1..20	1..20	no	a_1 scalar	-
Spider Chart	3..10	1..20	yes	a_1 scalar	-

Pattern

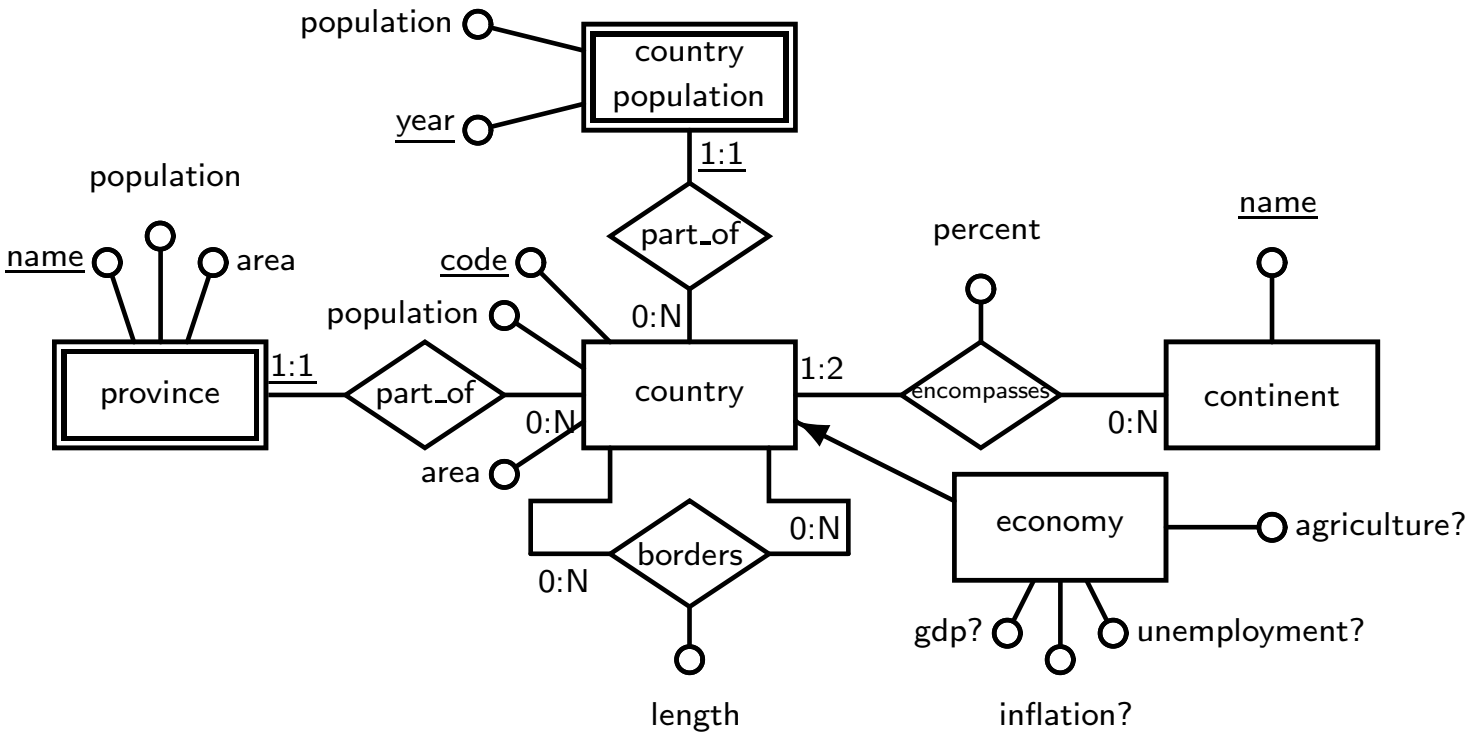


Match with the key of an entity, plus one or more attributes

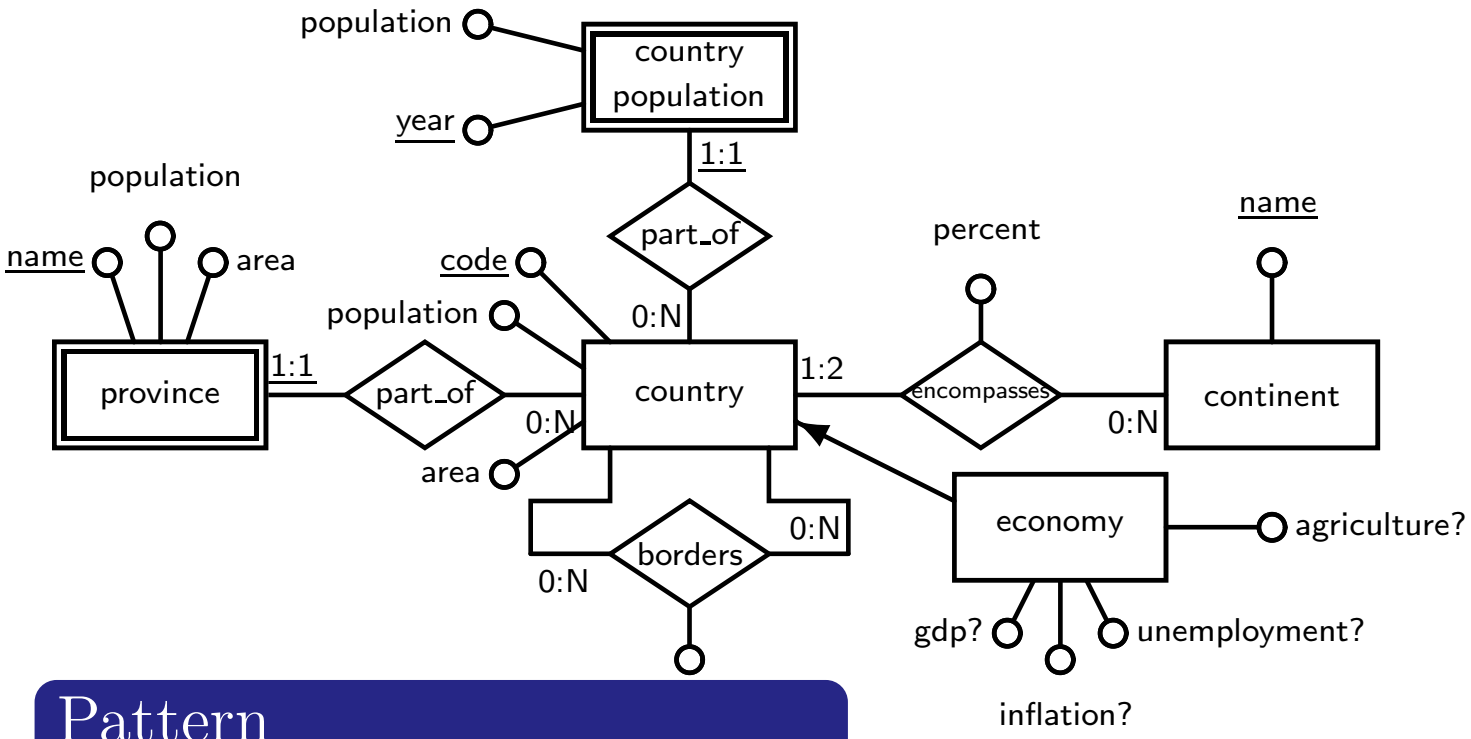
Instantiation



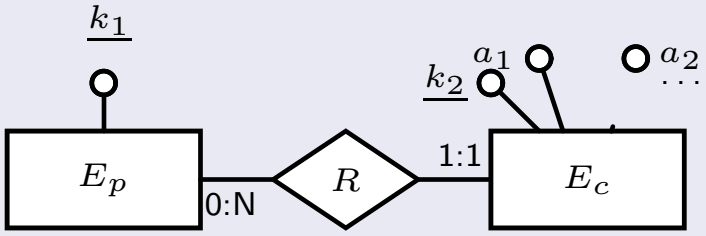
Visualisation Schema Pattern 3: One-Many Relationship



Visualisation Schema Pattern 3: One-Many Relationship

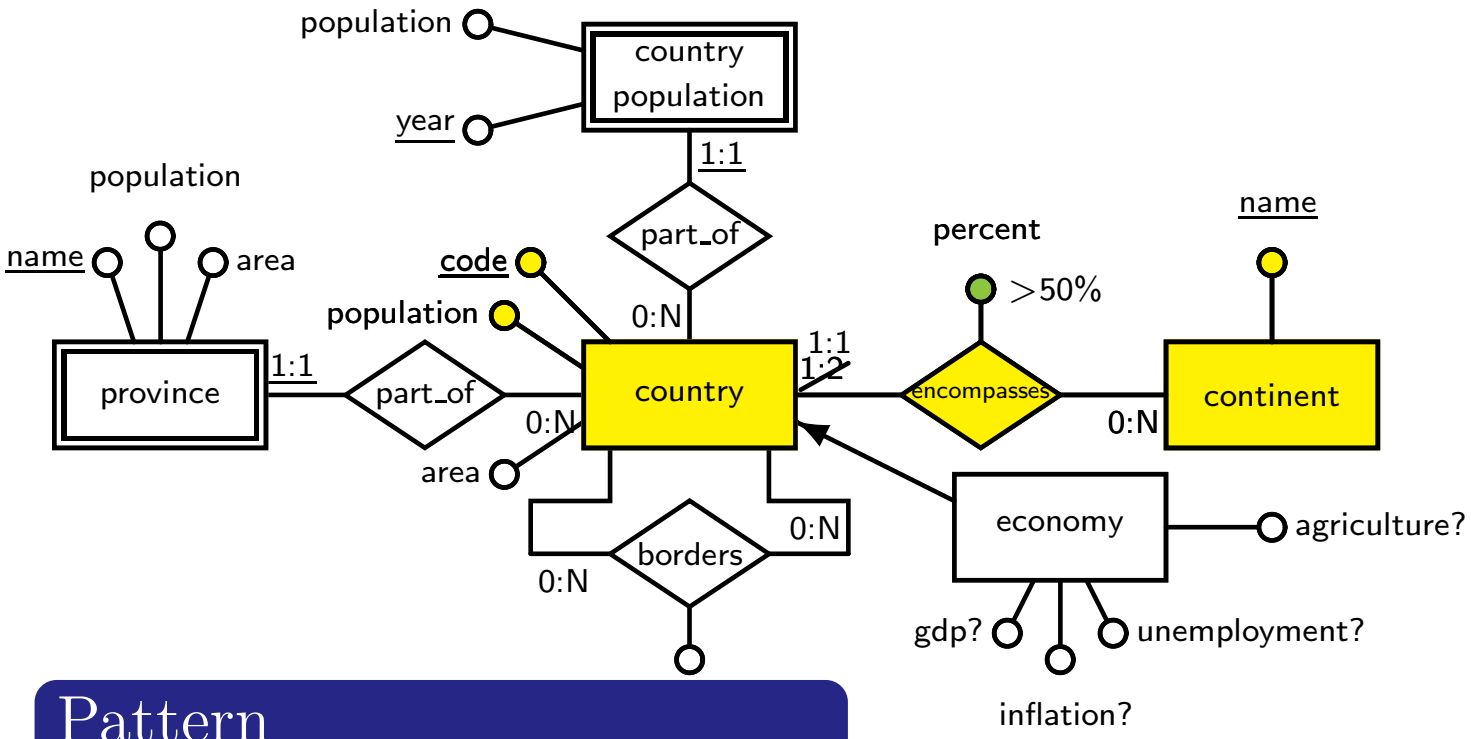


Pattern

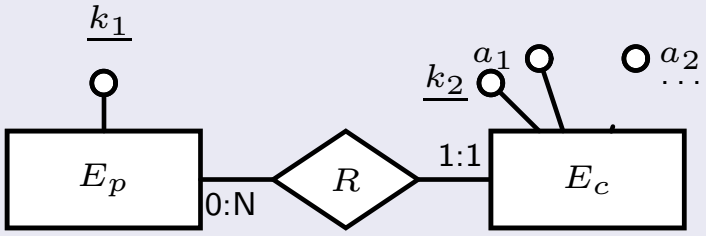


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 3: One-Many Relationship

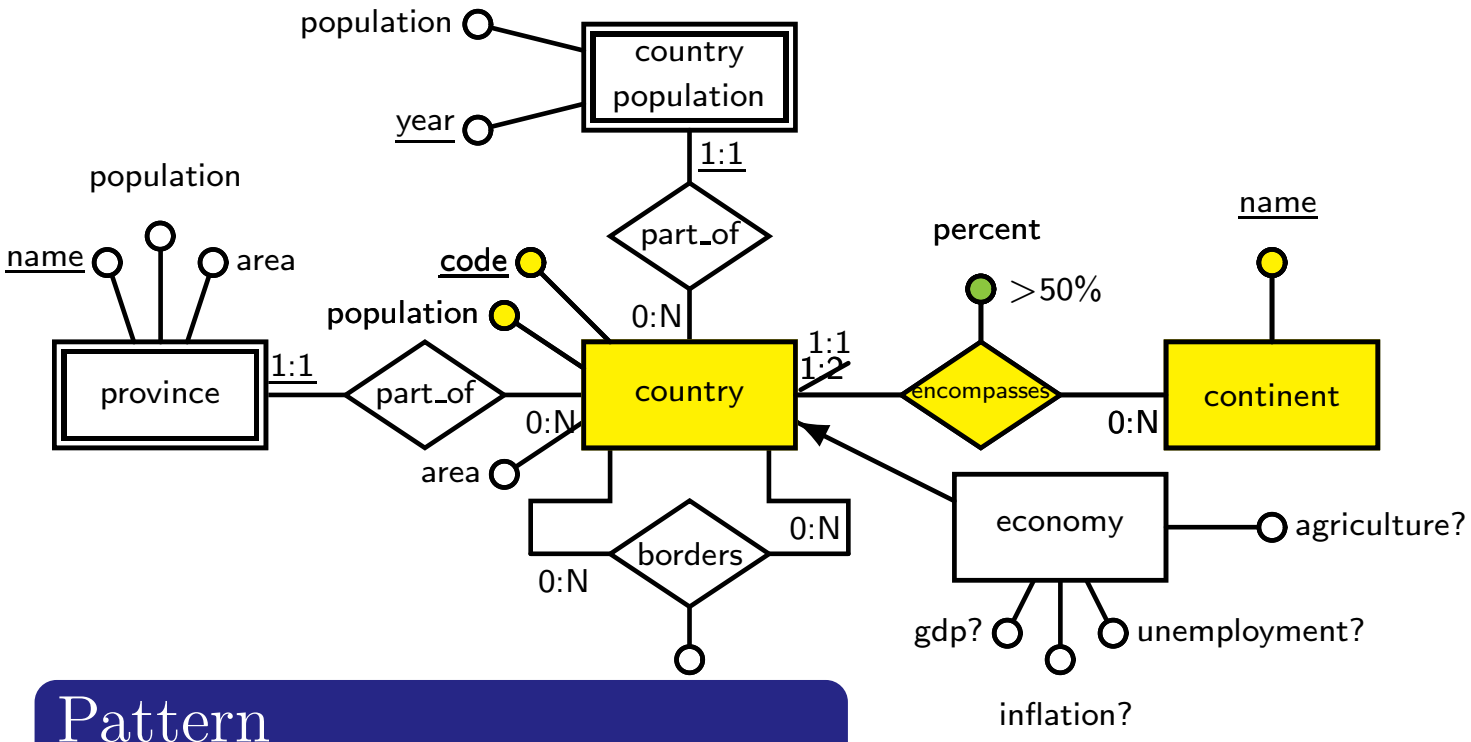


Pattern

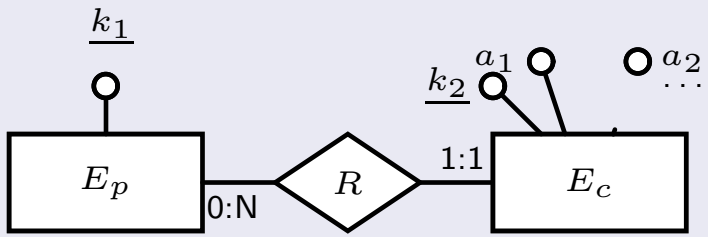


Match with the key of an entity, plus one or more attributes

Visualisation Schema Pattern 3: One-Many Relationship

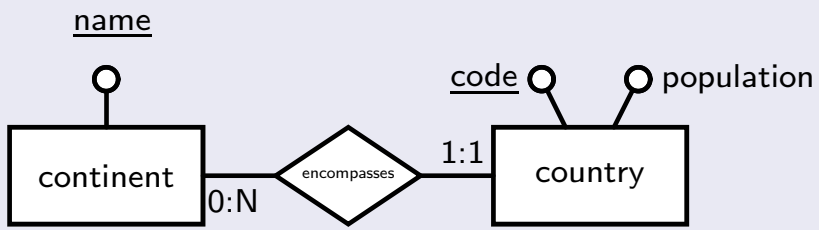


Pattern



Match with the key of an entity, plus one or more attributes

Instantiation

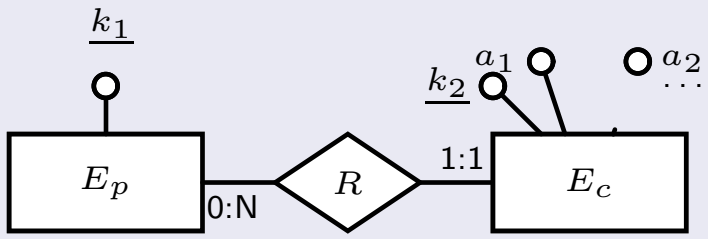


Visualisation Schema Pattern 3: One-Many Relationship

Visualisation Types

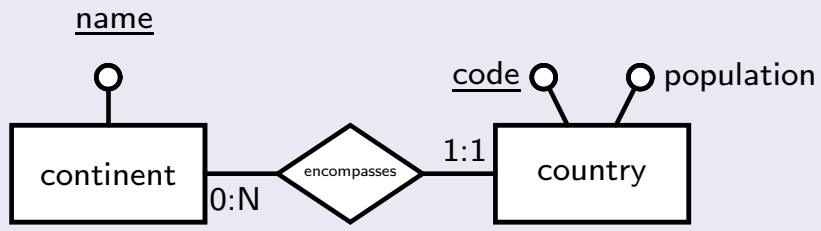
One-many relationships				
Name	$\lvert k_1 \rvert$	$\lvert k_2 \rvert$ per k_1	mandatory	optional
Tree Map	1..100	1..100	a_1 scalar	a_2 colour
Hierarchy Tree	1..100	1..100	-	a_1 colour
Circle Packing	1..100	1..100	a_1 scalar	a_2 colour

Pattern

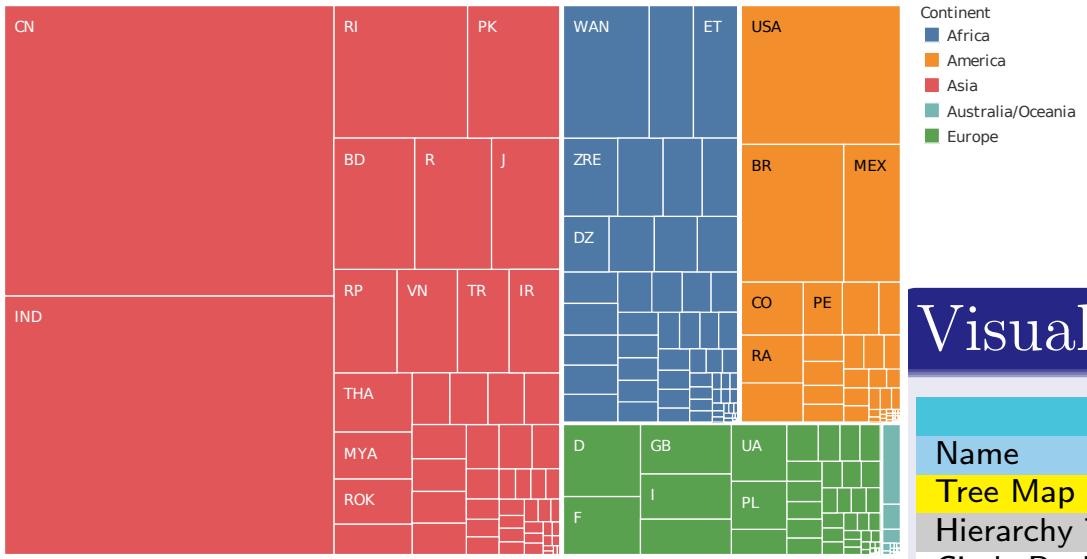


Match with the key of an entity, plus one or more attributes

Instantiation



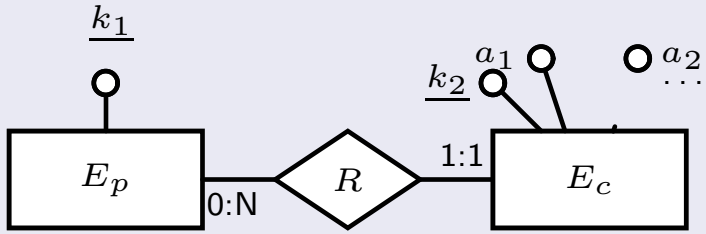
Visualisation Schema Pattern 3: One-Many Relationship



Visualisation Types

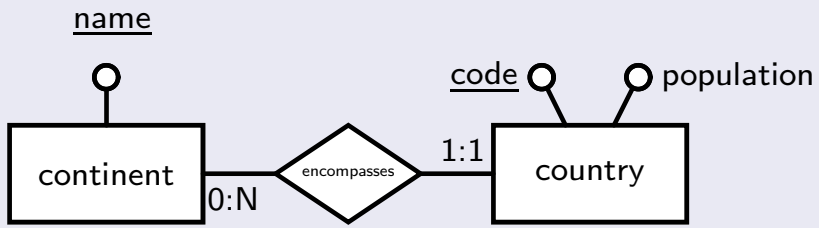
One-many relationships				
Name	$\|k_1\|$	$\|k_2\|$ per k_1	mandatory	optional
Tree Map	1..100	1..100	a_1 scalar	a_2 colour
Hierarchy Tree	1..100	1..100	-	a_1 colour
Circle Packing	1..100	1..100	a_1 scalar	a_2 colour

Pattern

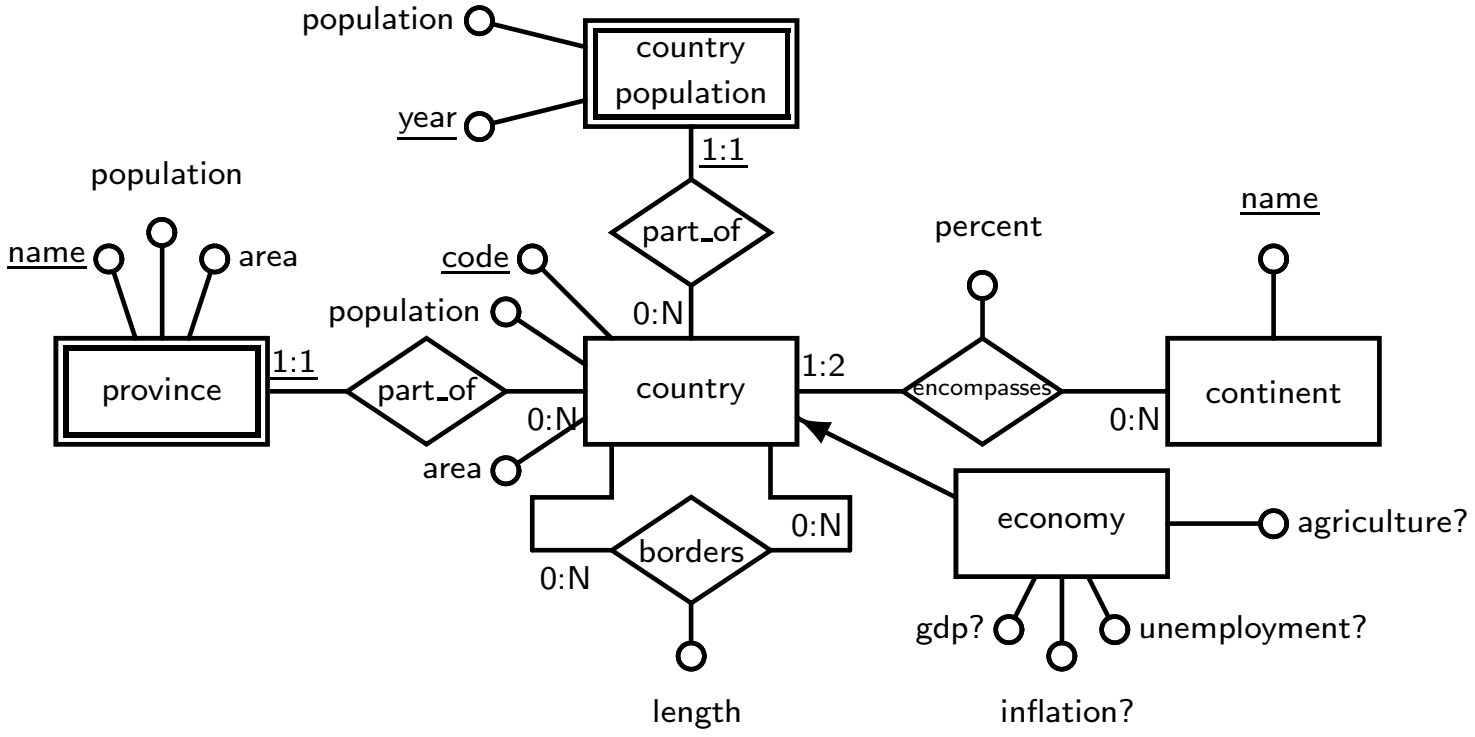


Match with the key of an entity, plus one or more attributes

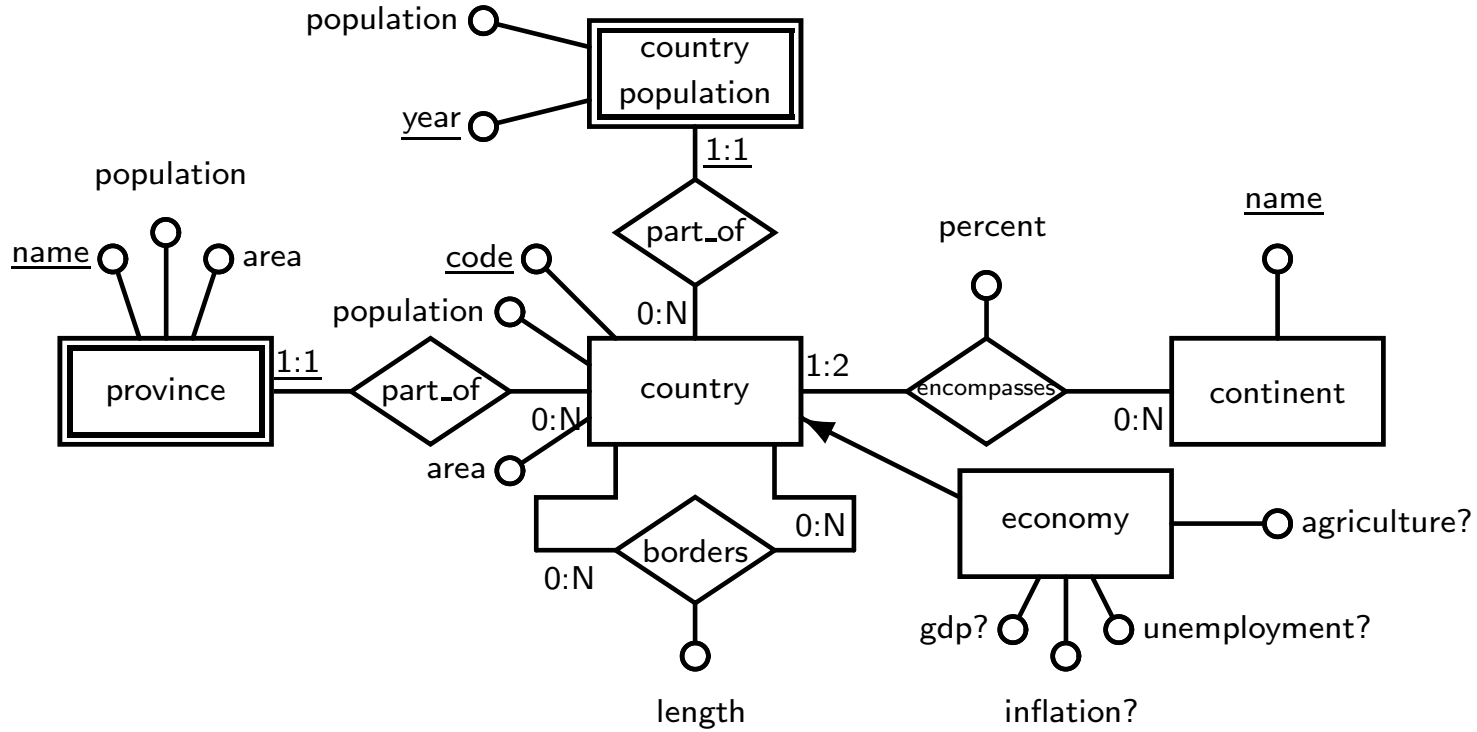
Instantiation



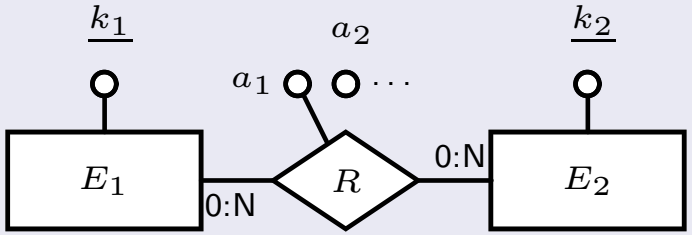
Visualisation Schema Pattern 4: Many-Many Relationship



Visualisation Schema Pattern 4: Many-Many Relationship

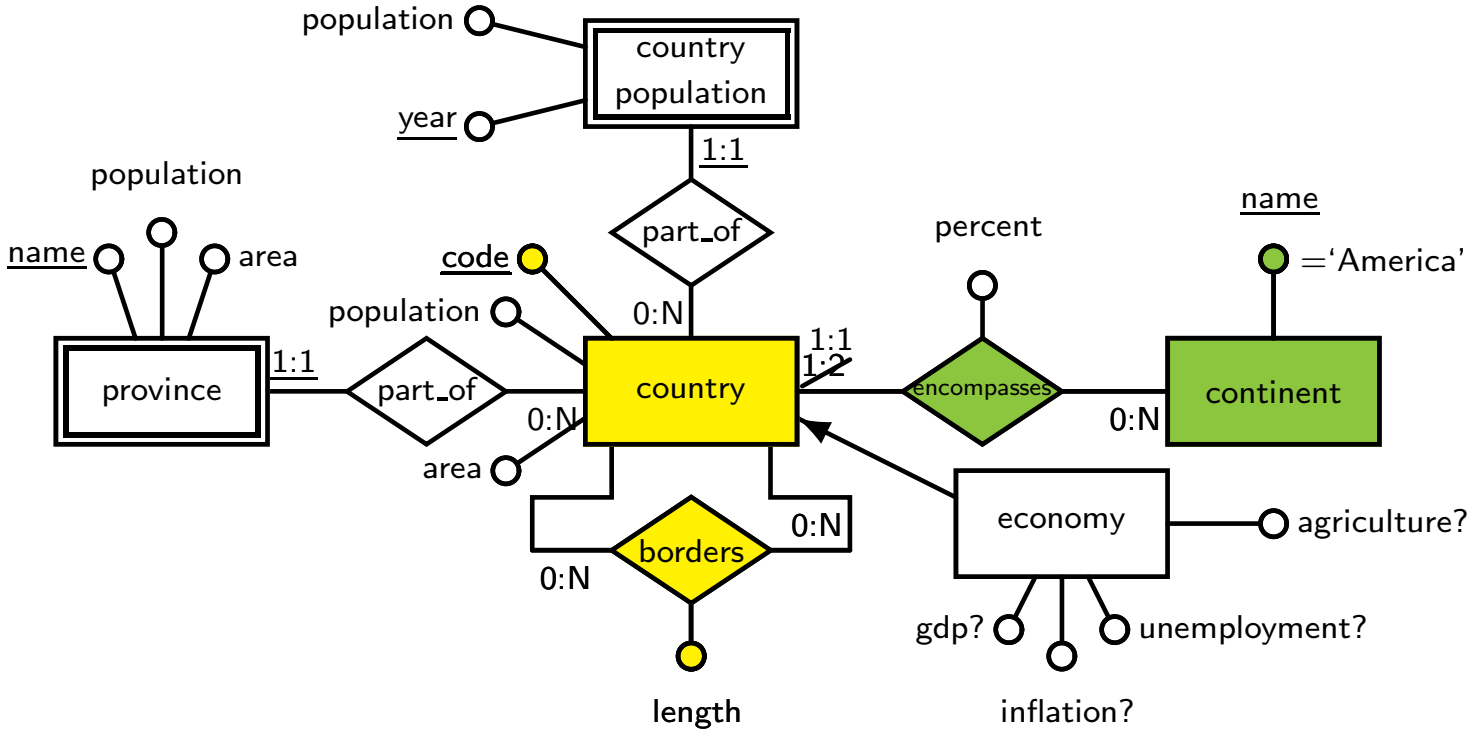


Pattern

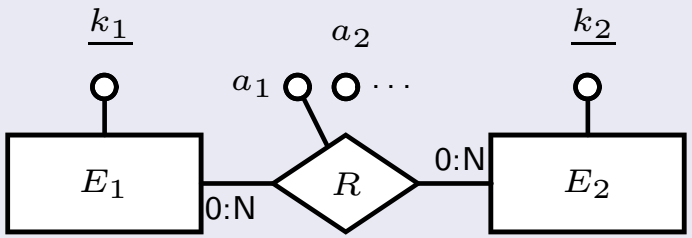


Match a (possibly reflexive) many-many relationship

Visualisation Schema Pattern 4: Many-Many Relationship

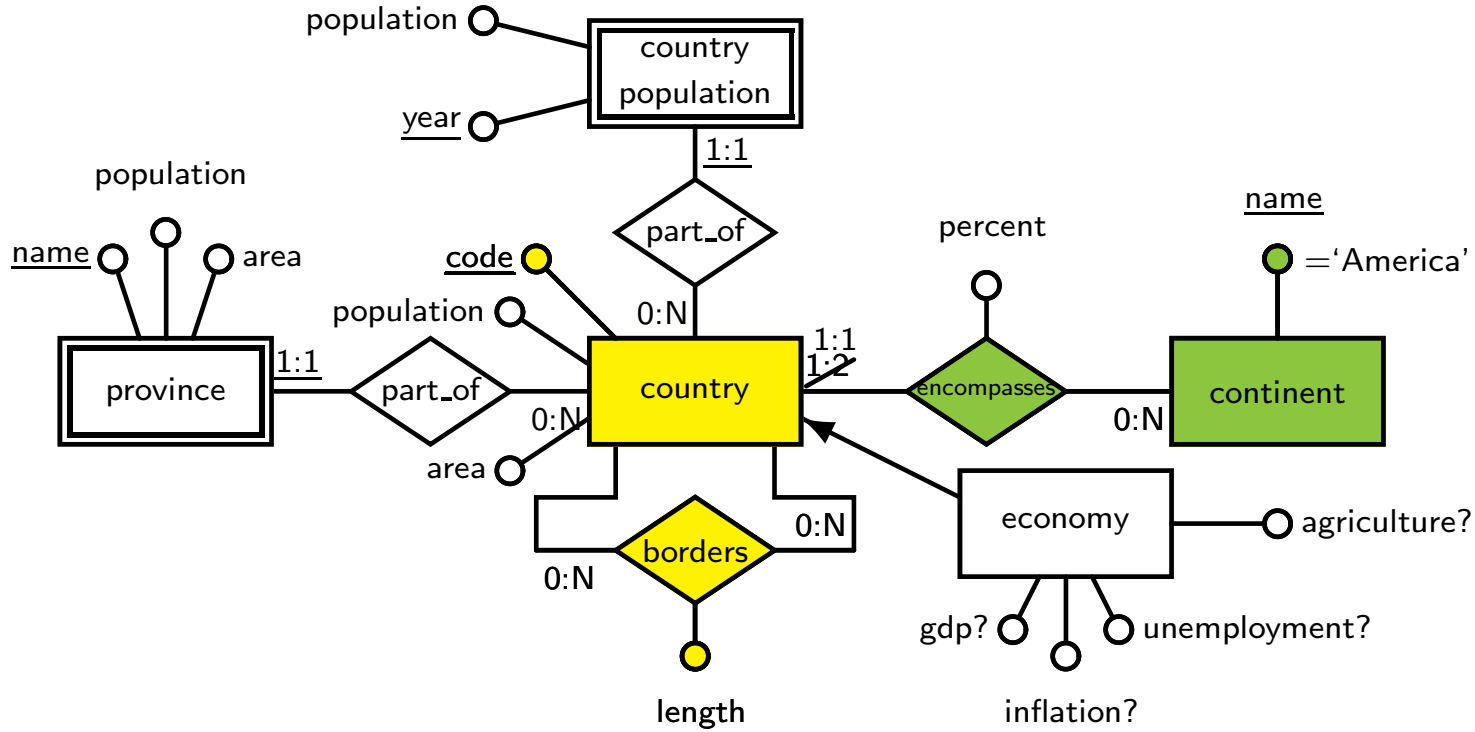


Pattern

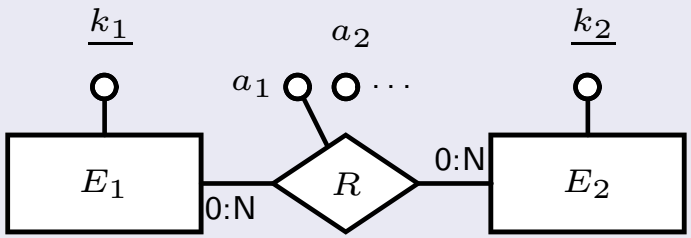


Match a (possibly reflexive) many-many relationship

Visualisation Schema Pattern 4: Many-Many Relationship

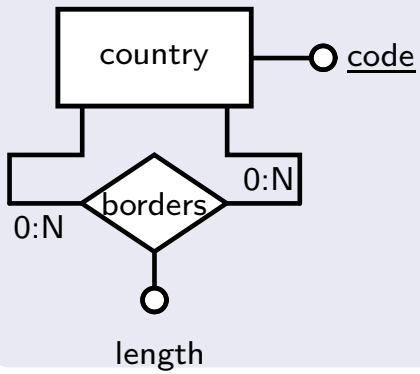


Pattern



Match a (possibly reflexive) many-many relationship

Instantiation

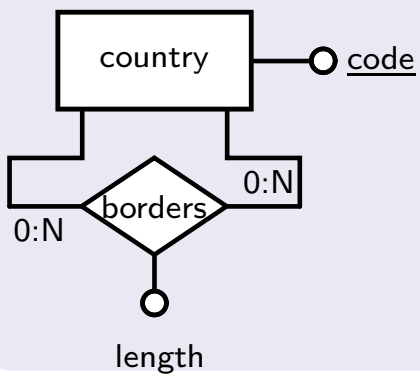


Visualisation Schema Pattern 4: Many-Many Relationship

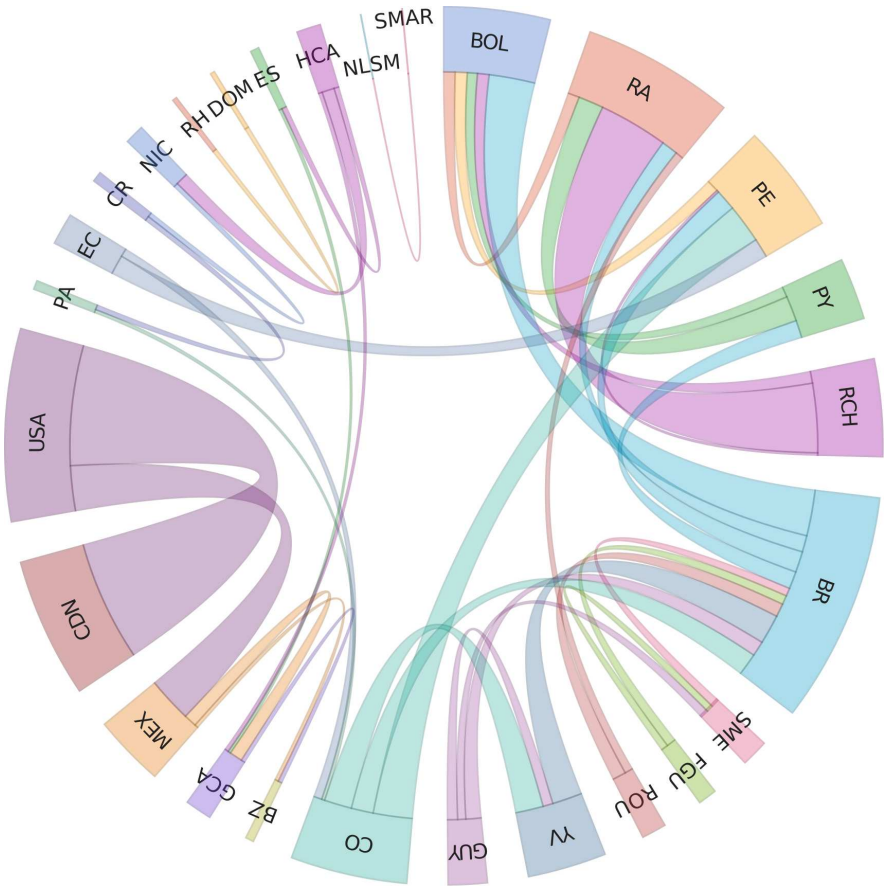
Visualisation Types

Many-many relationships					
Name	$\|k_1\|$	$\|k_2\|$	reflexive	mandatory	optional
Sankey	1..20	1..20	no	a_1 scalar	a_2 colour
Chord	1..100	1..100	yes	a_1 scalar	a_2 colour

Instantiation



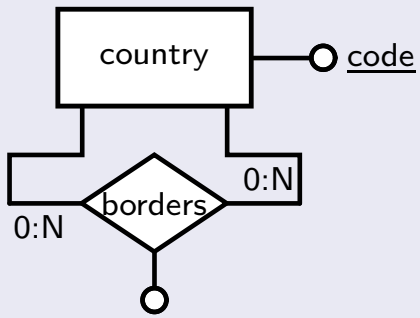
Visualisation Schema Pattern 4: Many-Many Relationship



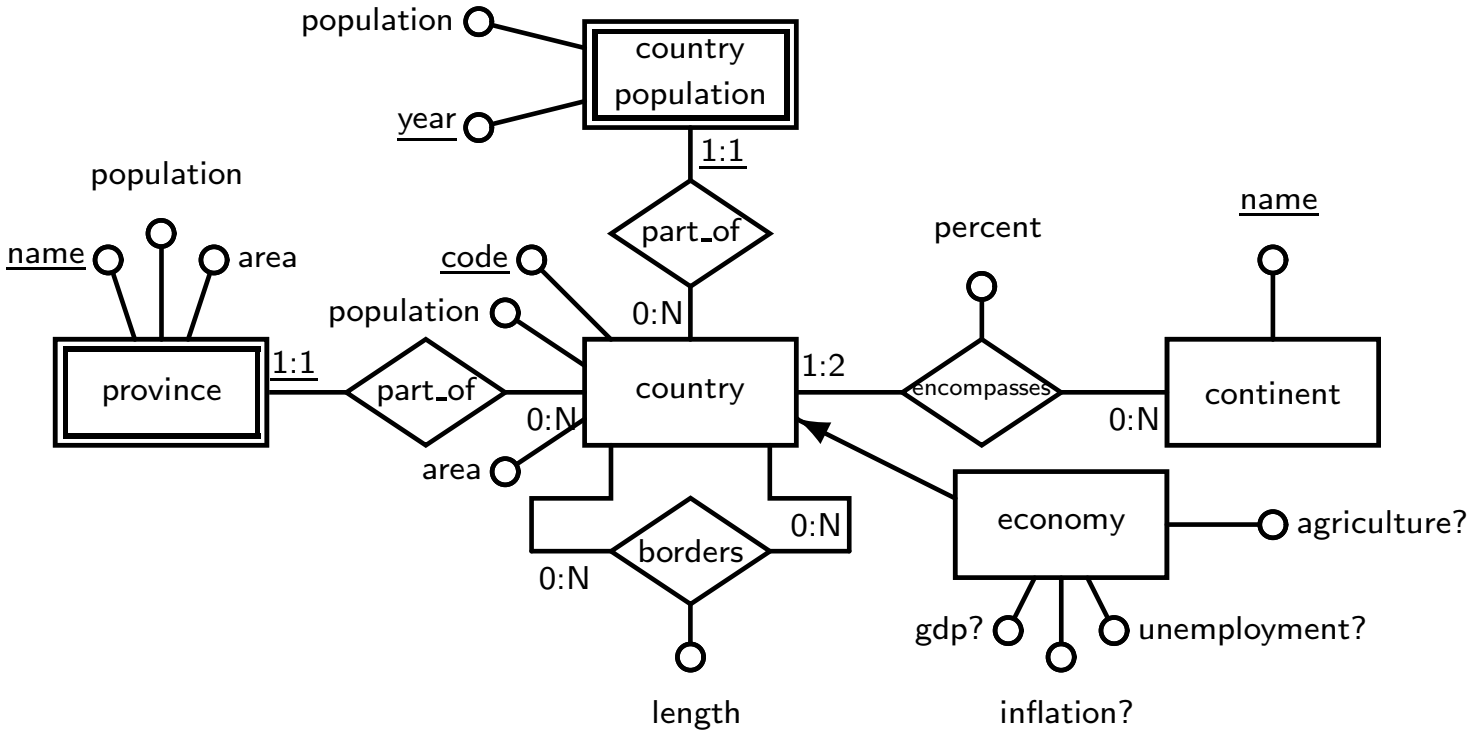
Visualisation Types

Many-many relationships					
Name	$\|k_1\|$	$\|k_2\|$	reflexive	mandatory	optional
Sankey	1..20	1..20	no	a_1 scalar	a_2 colour
Chord	1..100	1..100	yes	a_1 scalar	a_2 colour

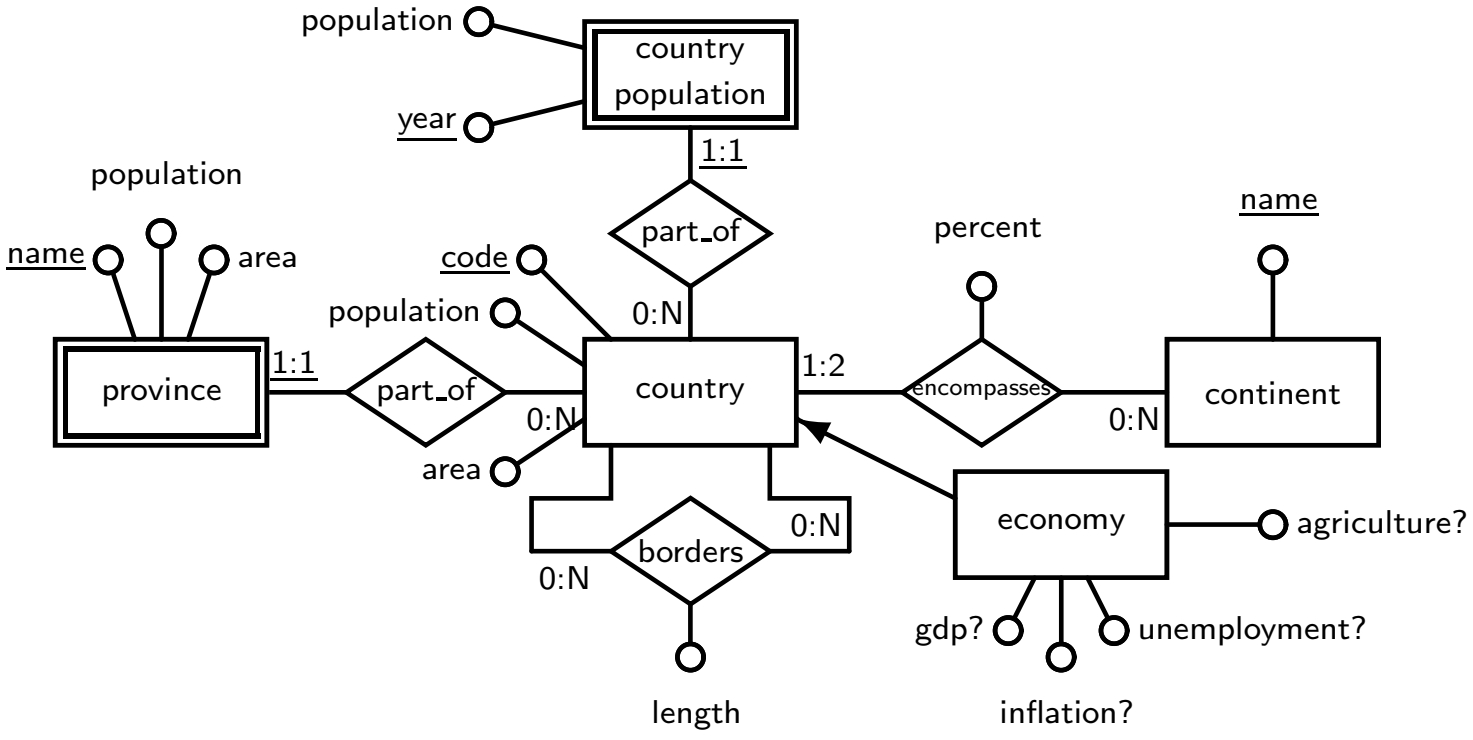
Instantiation



Transformations: Pivot



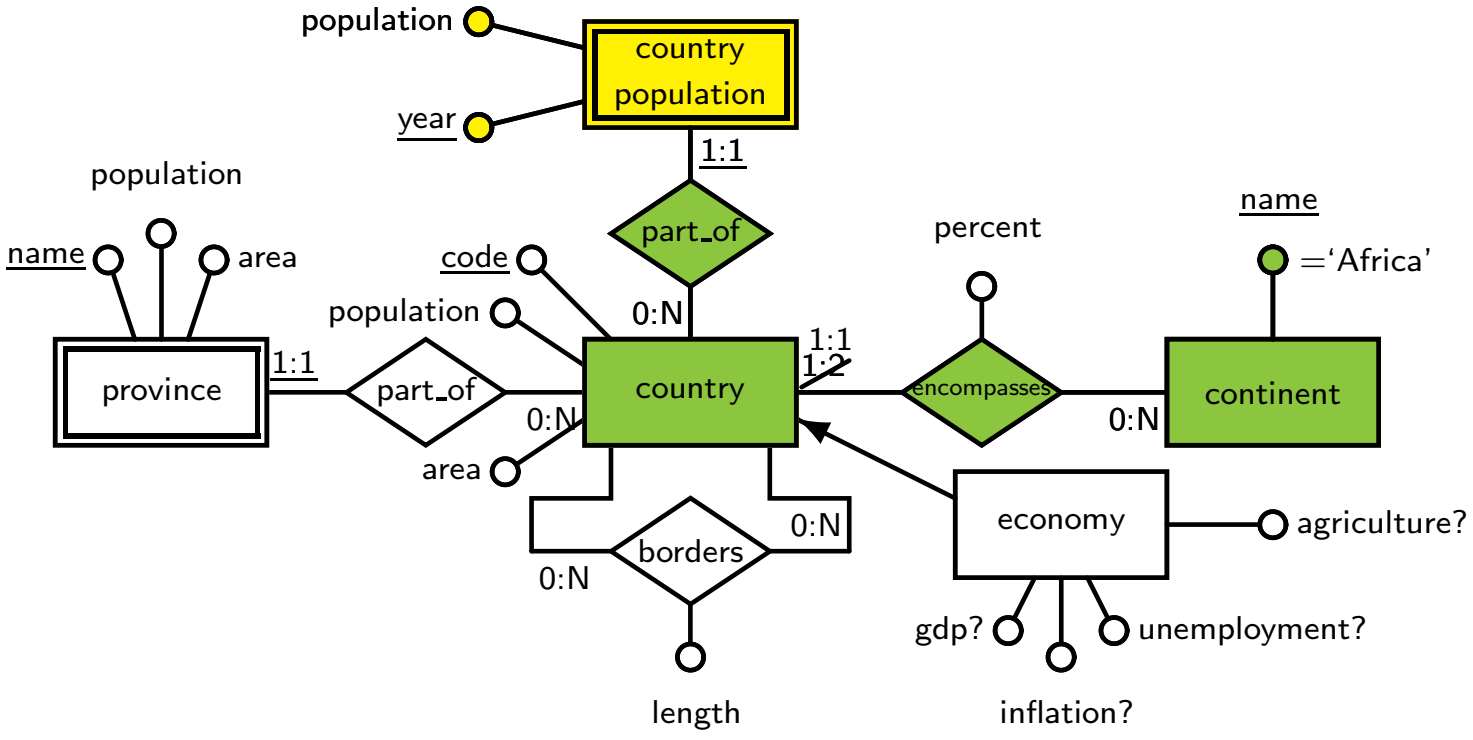
Transformations: Pivot



Transformation

Pivot $E(a, b)$ $\{v_1, \ldots, v_n\} \rightarrow$
$E_p(a\_v_1, \ldots, a\_v_n)$:-
$E(v_1, a\_v_1)$,
$\ldots$,
$E(v_n, a\_v_n)$.
Not information preserving

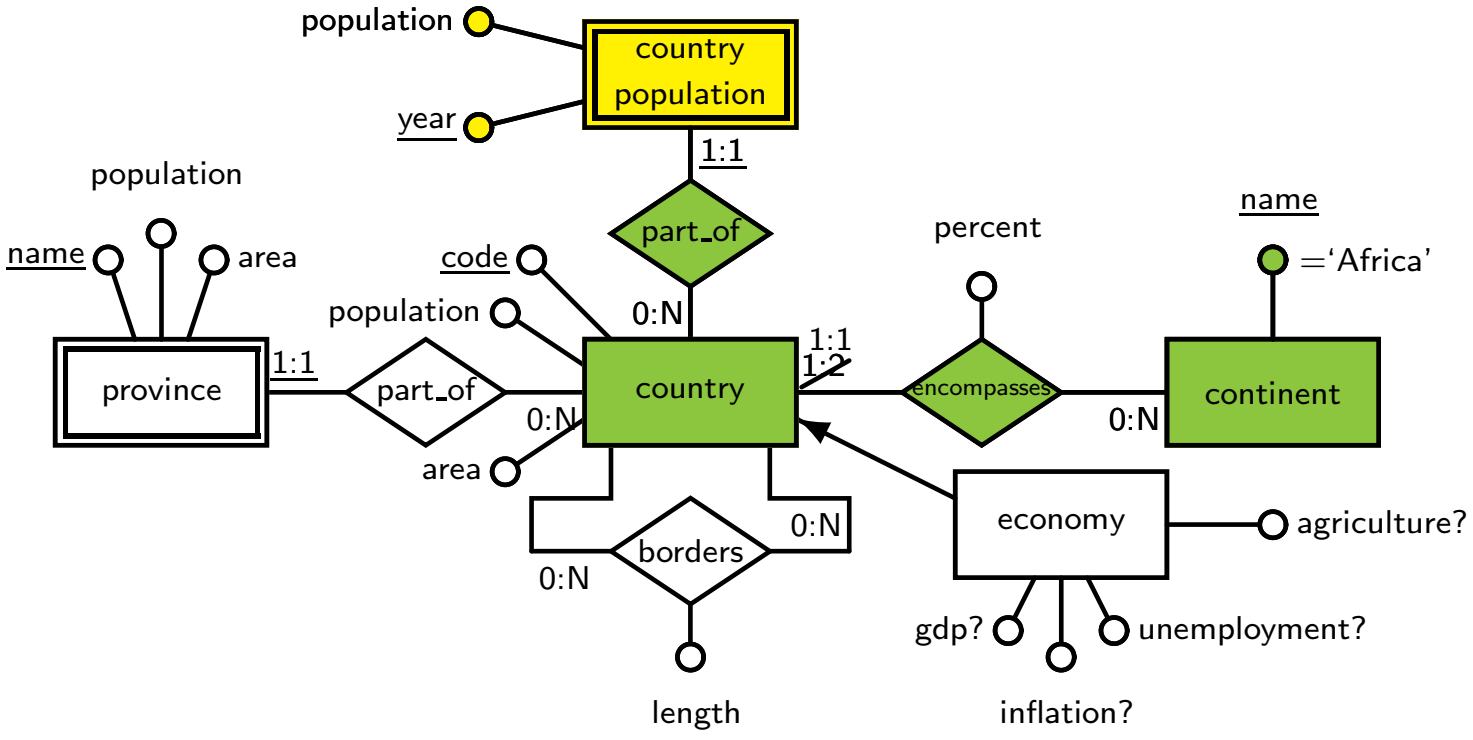
Transformations: Pivot



Transformation

Pivot $E(a, b)$ $\{v_1, \ldots, v_n\} \rightarrow$
$E_p(a\_v_1, \ldots, a\_v_n)$:-
$E(v_1, a\_v_1)$,
$\ldots$,
$E(v_n, a\_v_n)$.
Not information preserving

Transformations: Pivot

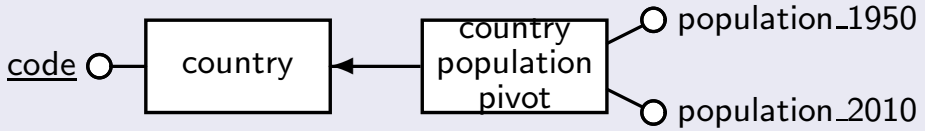


Transformation

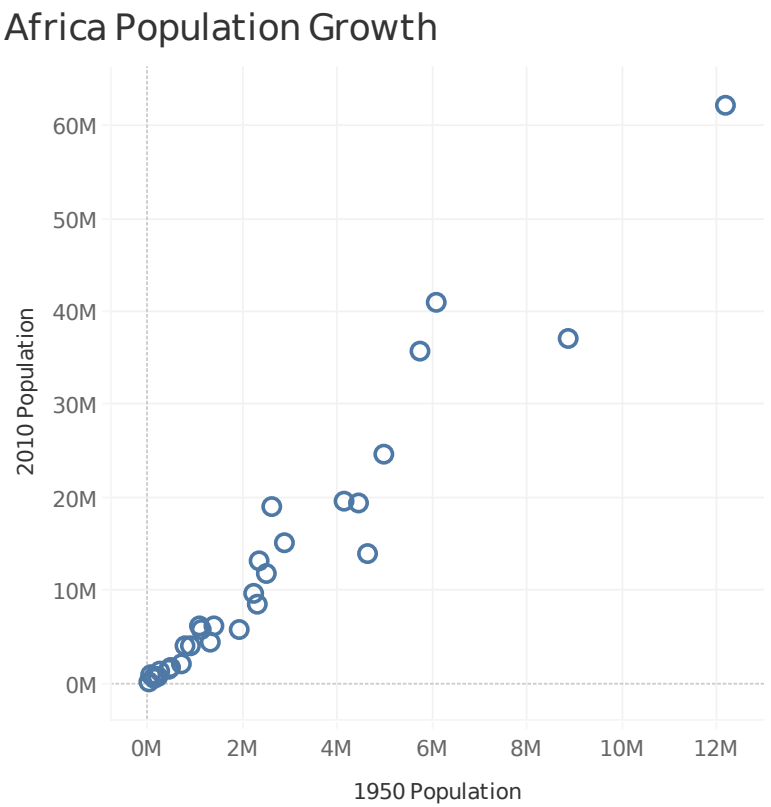
Pivot $E(a, b)$ $\{v_1, \ldots, v_n\} \rightarrow$
$E_p(a\_v_1, \ldots, a\_v_n)$:-
$E(v_1, a\_v_1)$,
$\ldots$,
$E(v_n, a\_v_n)$.
Not information preserving

Application to Schema

Pivot country_population(year, population) {1950, 2010}



Transformations: Pivot



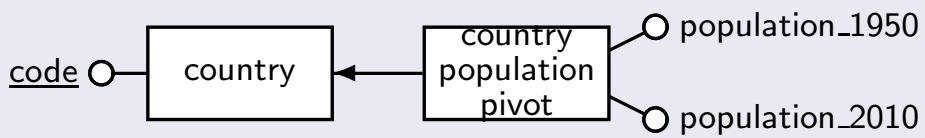
Transformation

Pivot $E(a, b)$ $\{v_1, \ldots, v_n\} \rightarrow$
$E_p(a\_v_1, \ldots, a\_v_n)$:-
$E(v_1, a\_v_1)$,
$\ldots$,
$E(v_n, a\_v_n)$.

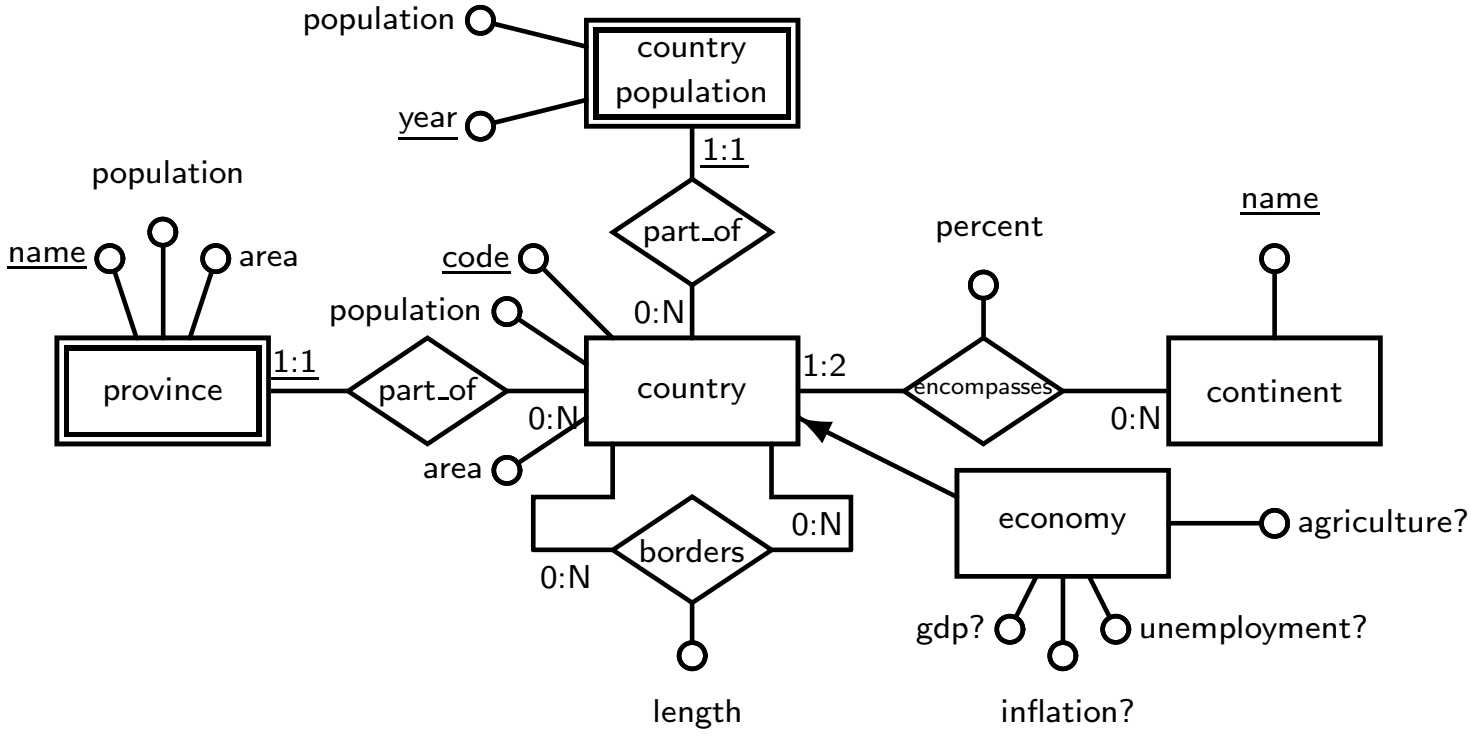
Not information preserving

Application to Schema

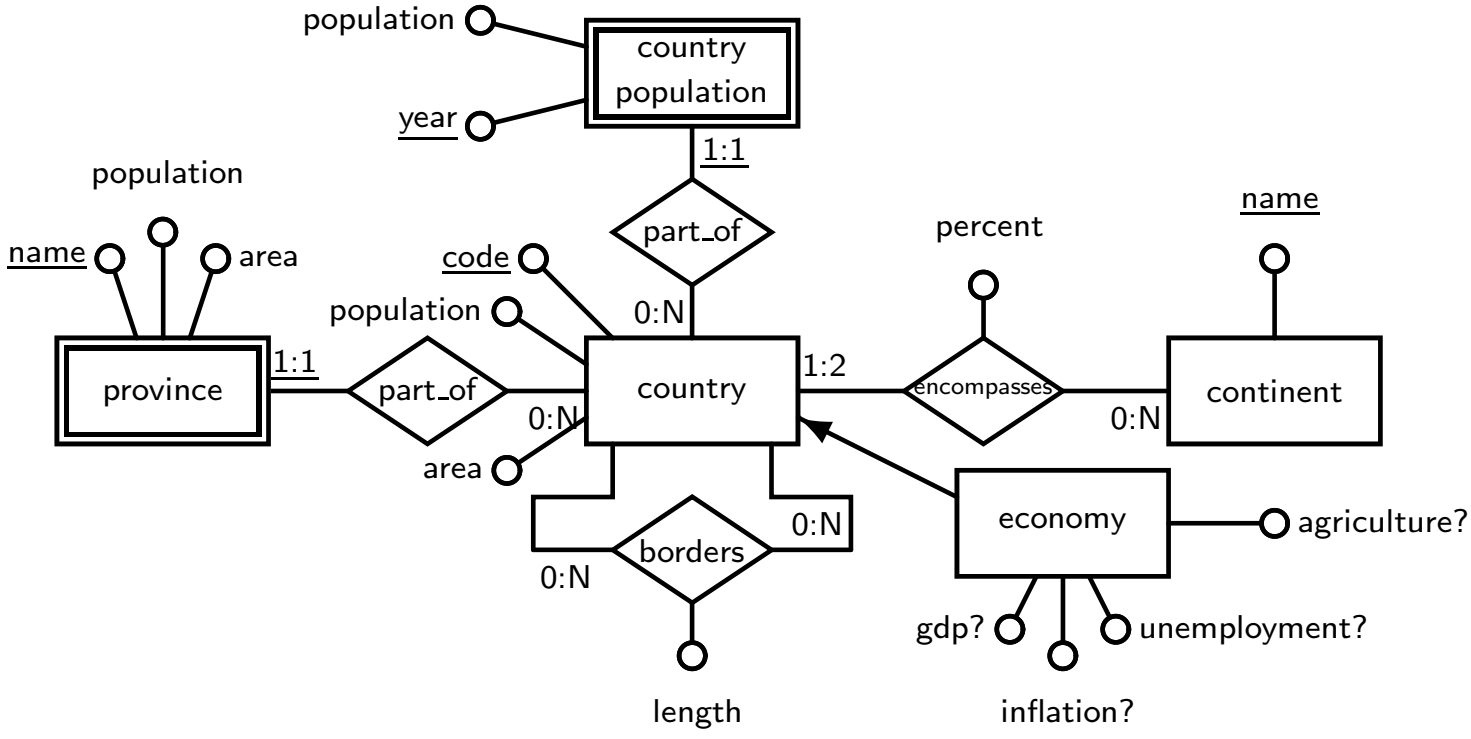
Pivot country_population(year, population) {1950, 2010}



Transformations: Denormalisation



Transformations: Denormalisation

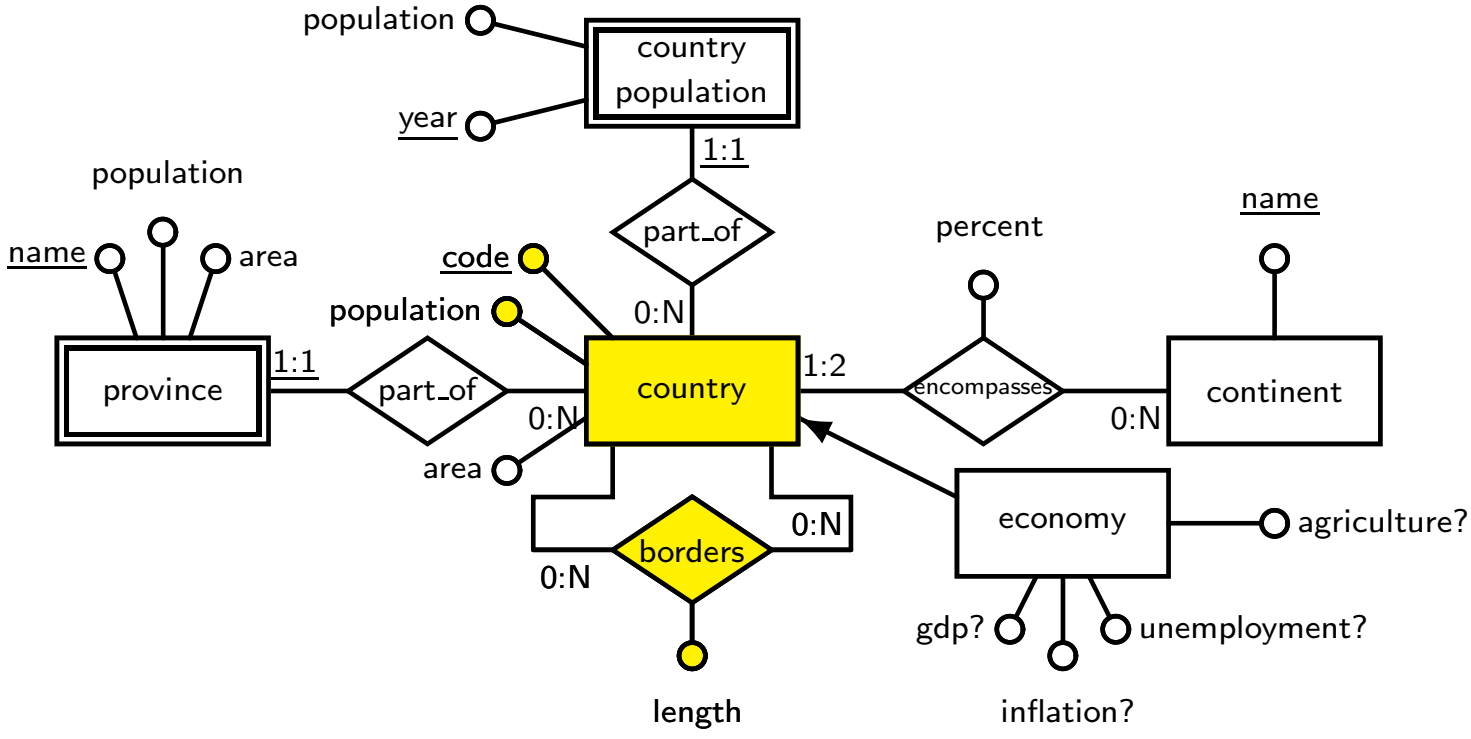


Transformation

Denormalise $E_2(k_2, x, \ldots)$
$R(k_1, k_2, a, \ldots) \rightarrow$
$ED_2(k_1, k_2, x, \ldots, a, \ldots)$:-
$E_2(k_2, x, \ldots)$,
$R(k_1, k_2, a, \ldots)$.

Information preserving

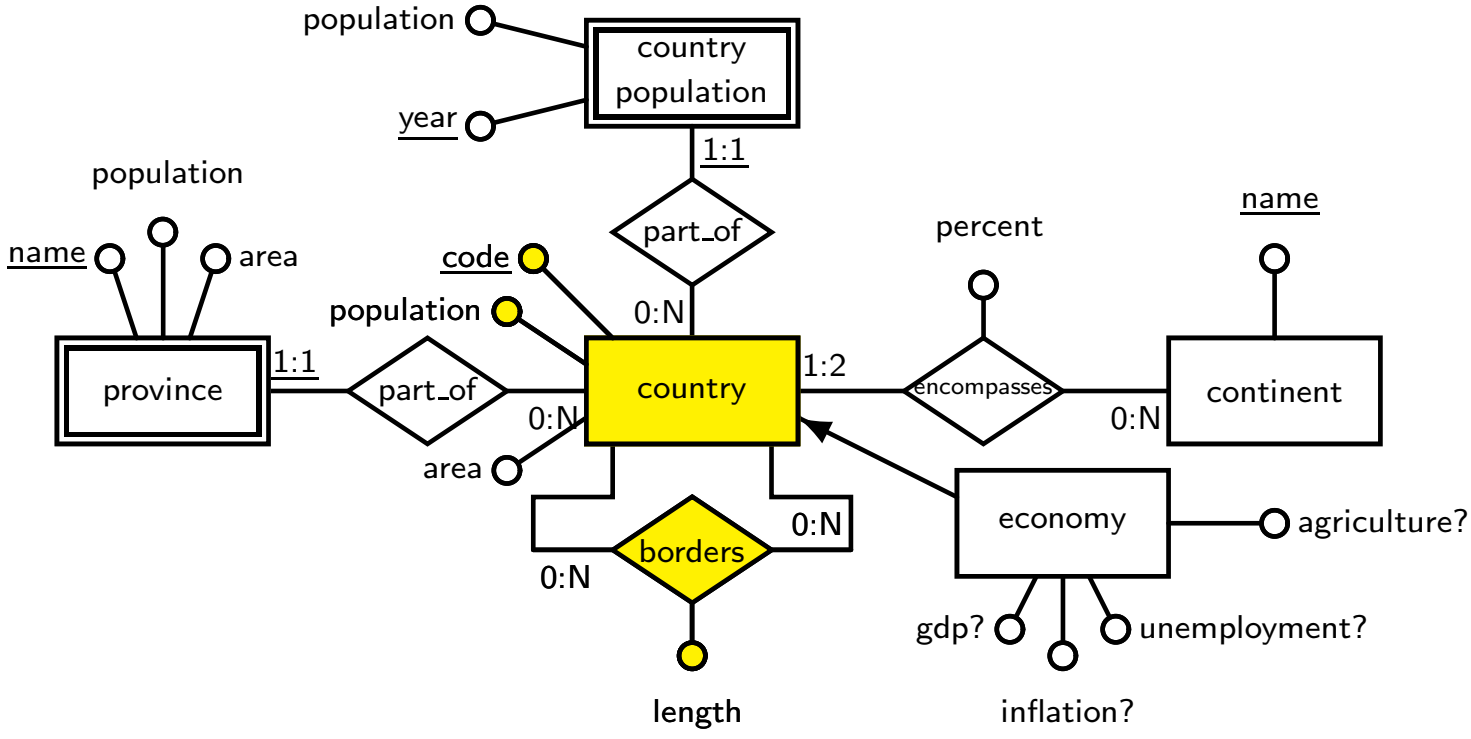
Transformations: Denormalisation



Transformation

Denormalise $E_2(k_2, x, \ldots)$
$R(k_1, k_2, a, \ldots) \rightarrow$
$ED_2(k_1, k_2, x, \ldots, a, \ldots)$:-
$E_2(k_2, x, \ldots)$,
$R(k_1, k_2, a, \ldots)$.
Information preserving

Transformations: Denormalisation

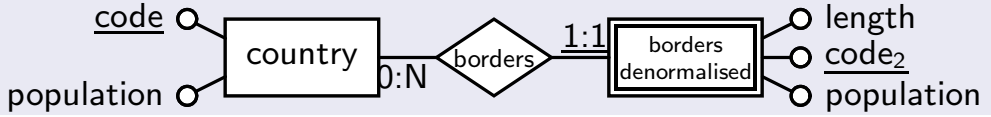


Transformation

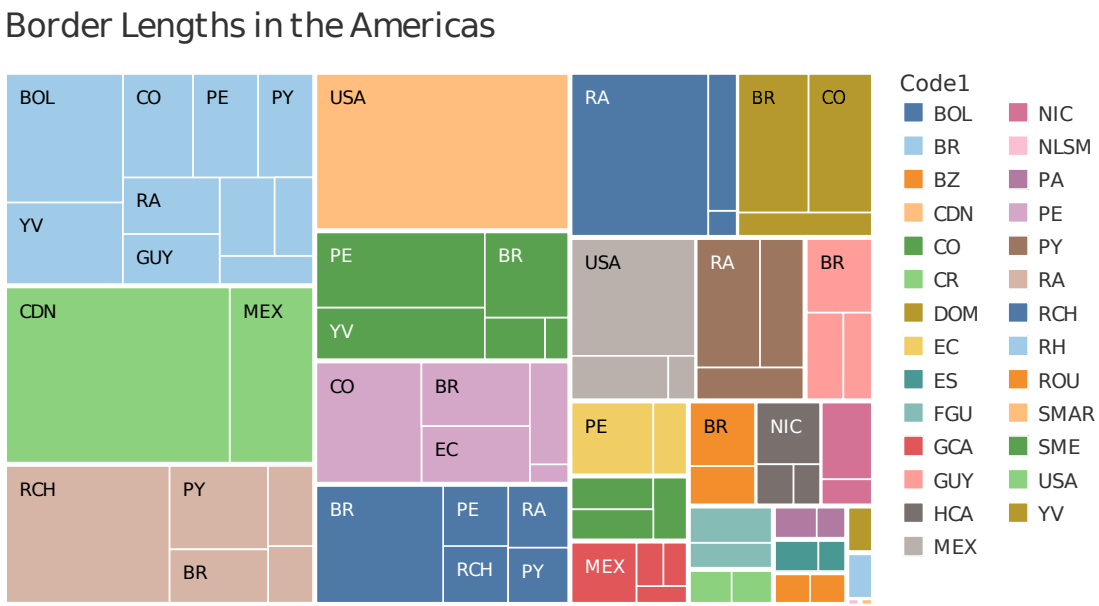
Denormalise $E_2(k_2, x, \ldots)$
$R(k_1, k_2, a, \ldots) \rightarrow$
$ED_2(k_1, k_2, x, \ldots, a, \ldots)$:-
$E_2(k_2, x, \ldots)$,
$R(k_1, k_2, a, \ldots)$.
Information preserving

Application to Schema

Denormalise country(code, population) borders($code_1$, $code_2$, length)



Transformations: Denormalisation



Transformation

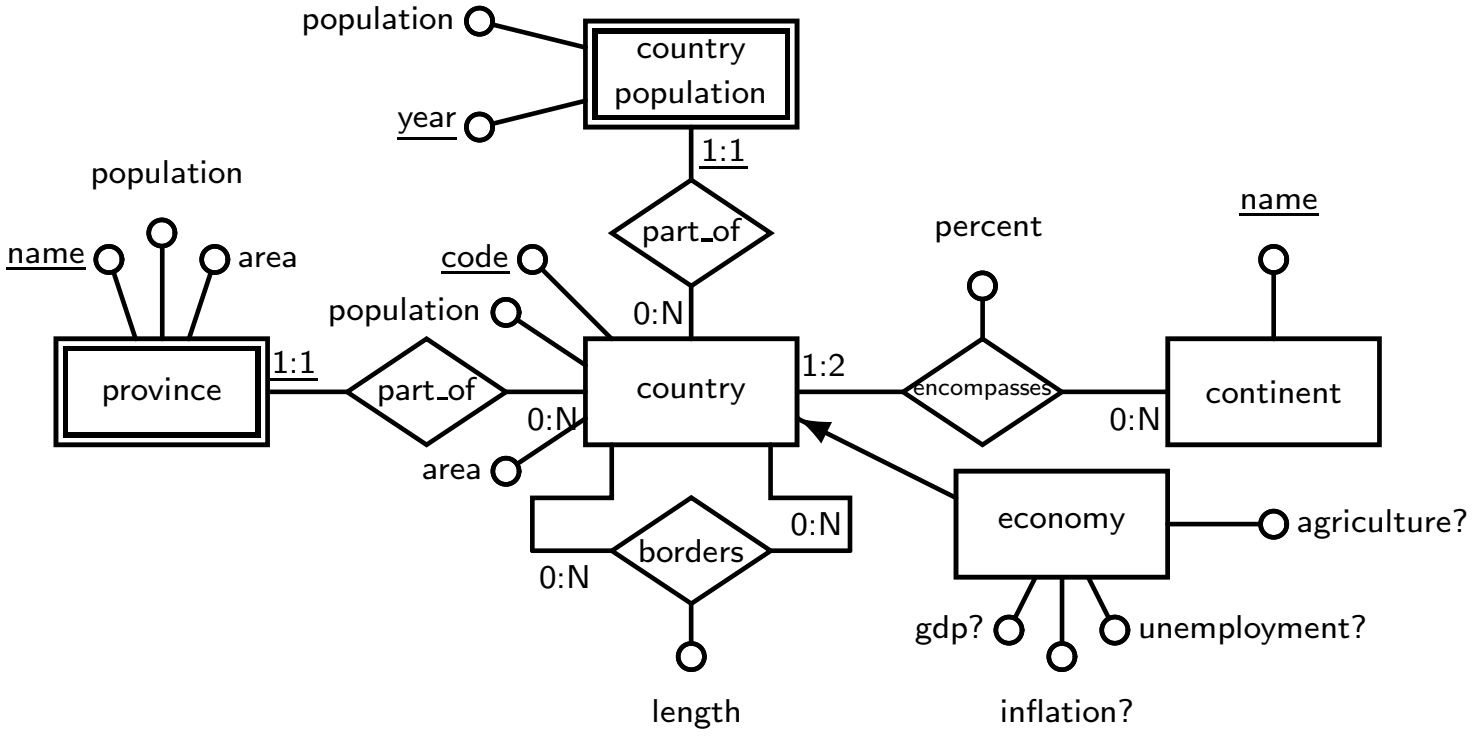
Denormalise $E_2(k_2, x, \ldots)$
$R(k_1, k_2, a, \ldots) \rightarrow$
$ED_2(k_1, k_2, x, \ldots, a, \ldots)$:-
$E_2(k_2, x, \ldots)$,
$R(k_1, k_2, a, \ldots)$.

Information preserving

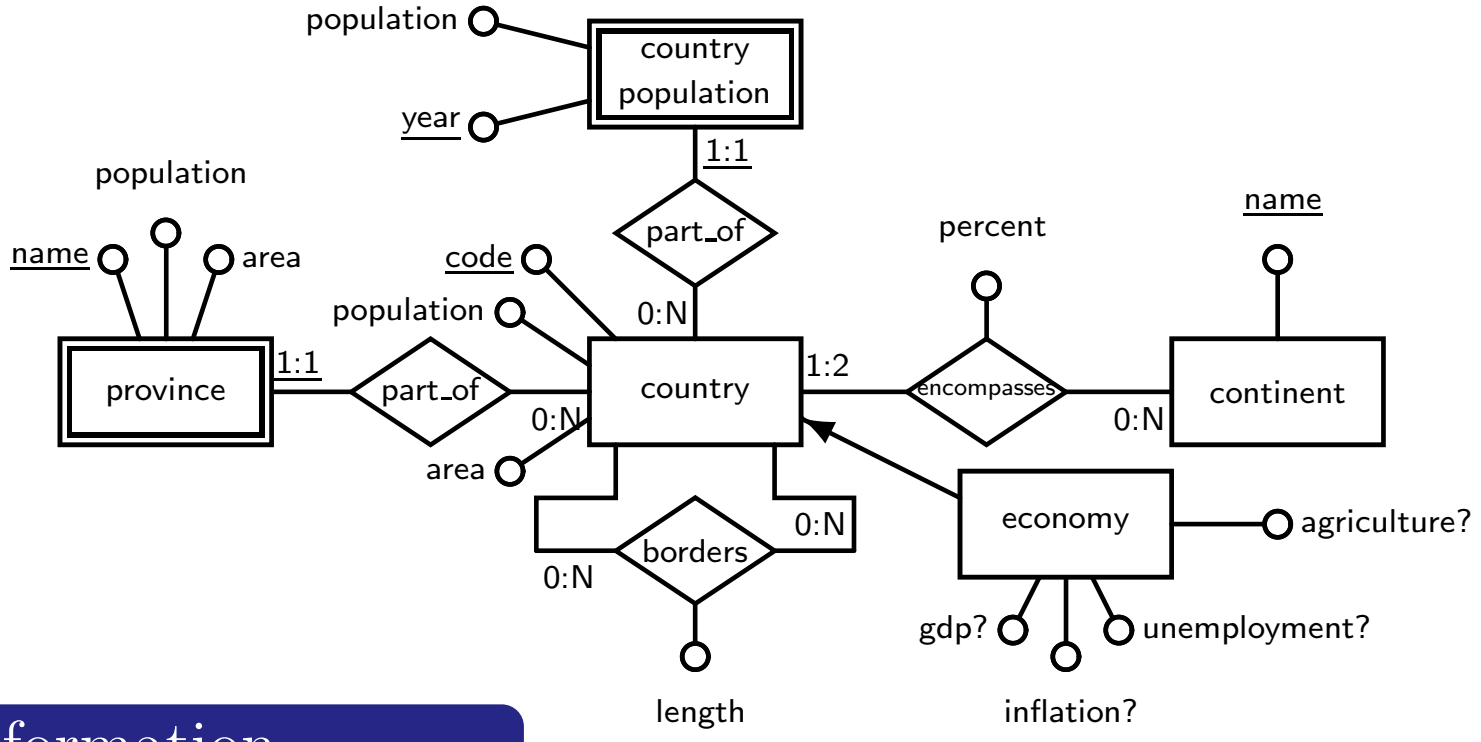
Application to Schema

Denormalise country(code, population) borders($code_1$, $code_2$, length)

Transformations: Attribute Specialisation



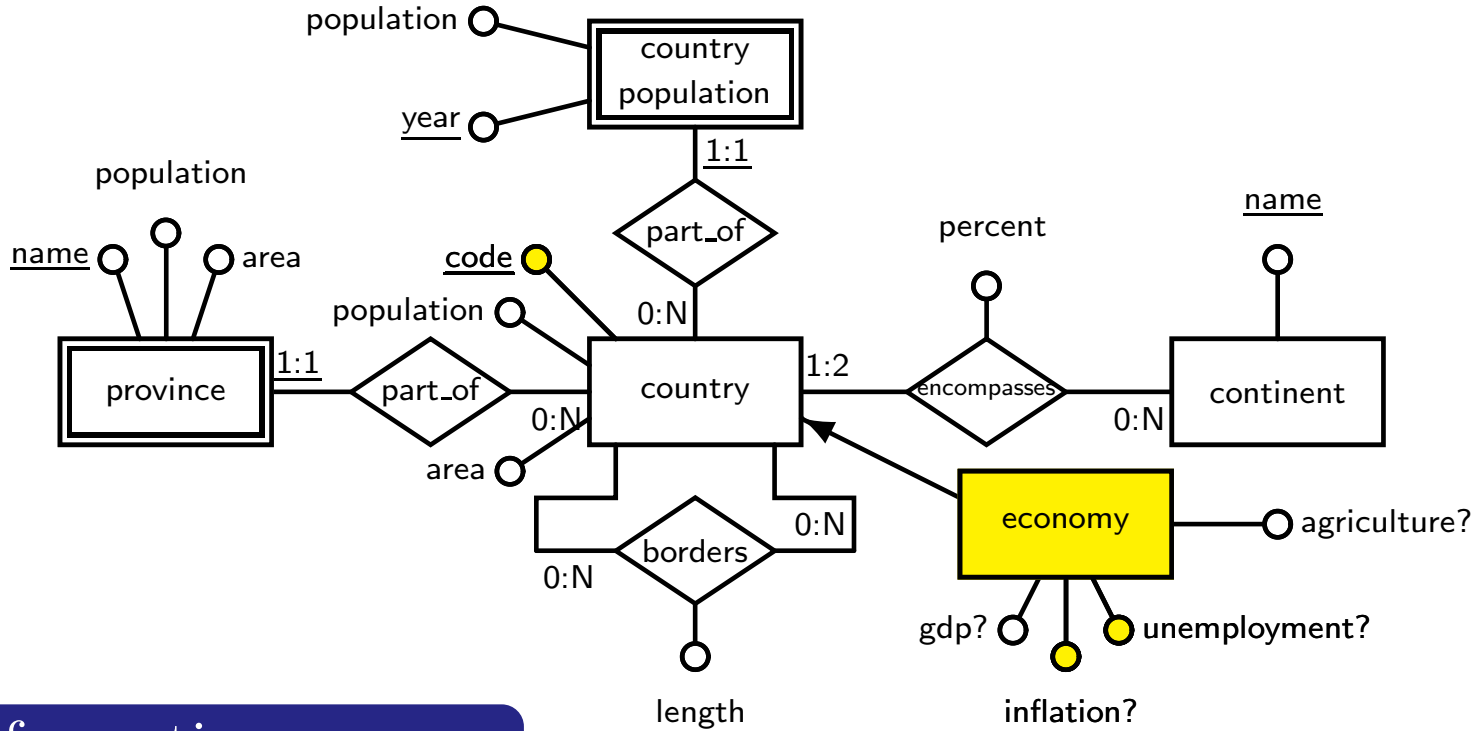
Transformations: Attribute Specialisation



Transformation

Mandatory $E(k, a_1, \ldots, a_n) \rightarrow$
$EM(k, a_1, \ldots, a_n)$:-
$E(k, a_1, \ldots, a_n)$,
$IsNotNull(a_1)$,
$\ldots$,
$IsNotNull(a_n)$.
Not information preserving

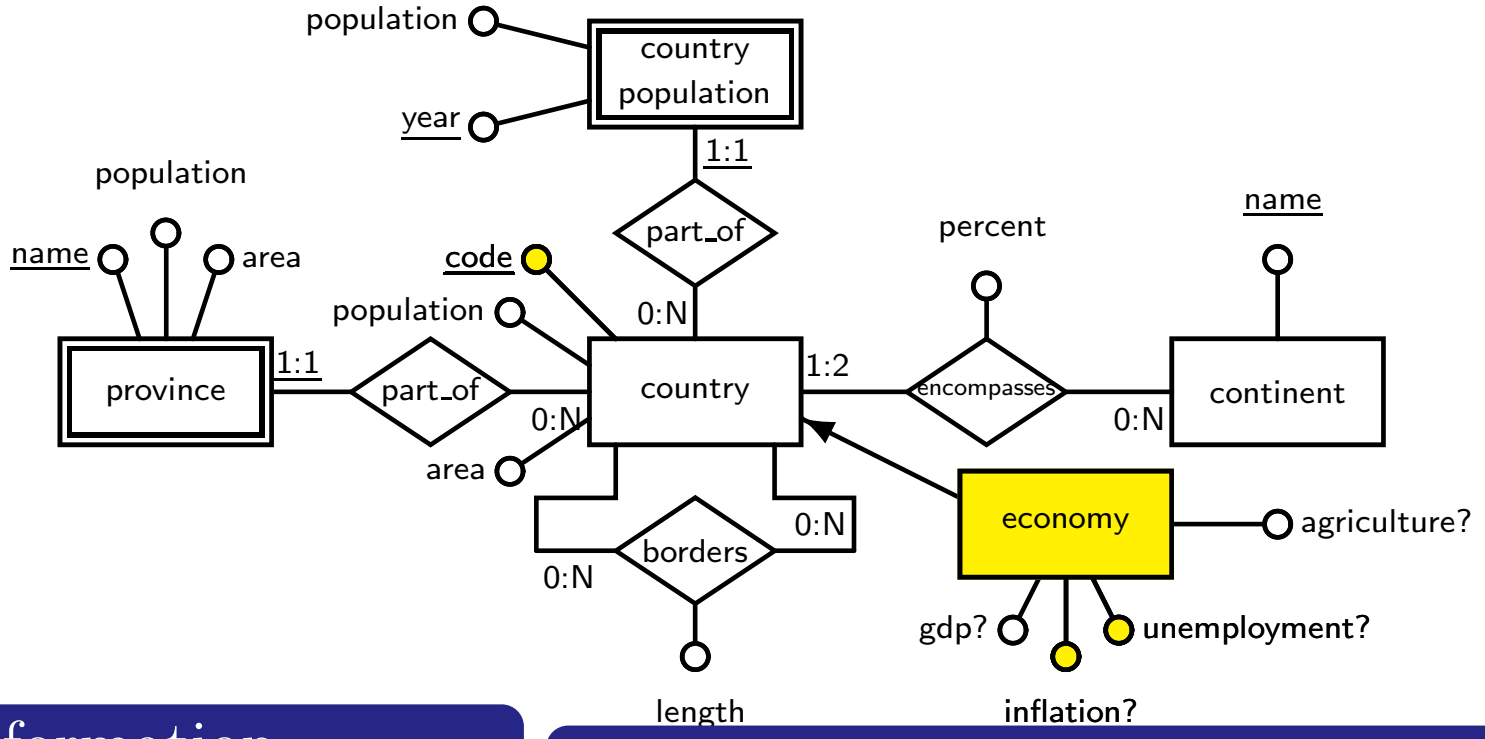
Transformations: Attribute Specialisation



Transformation

Mandatory $E(k, a_1, \ldots, a_n) \rightarrow$
$EM(k, a_1, \ldots, a_n)$:-
$E(k, a_1, \ldots, a_n)$,
$\text{IsNotNull}(a_1)$,
$\ldots$,
$\text{IsNotNull}(a_n)$.
Not information preserving

Transformations: Attribute Specialisation



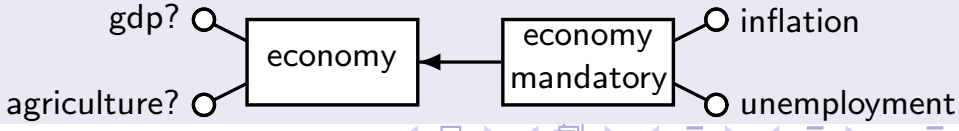
Transformation

Mandatory $E(k, a_1, \ldots, a_n) \rightarrow$
$EM(k, a_1, \ldots, a_n)$:-
$E(k, a_1, \ldots, a_n)$,
$\mathsf{IsNotNull}(a_1)$,
$\ldots$,
$\mathsf{IsNotNull}(a_n)$.

Not information preserving

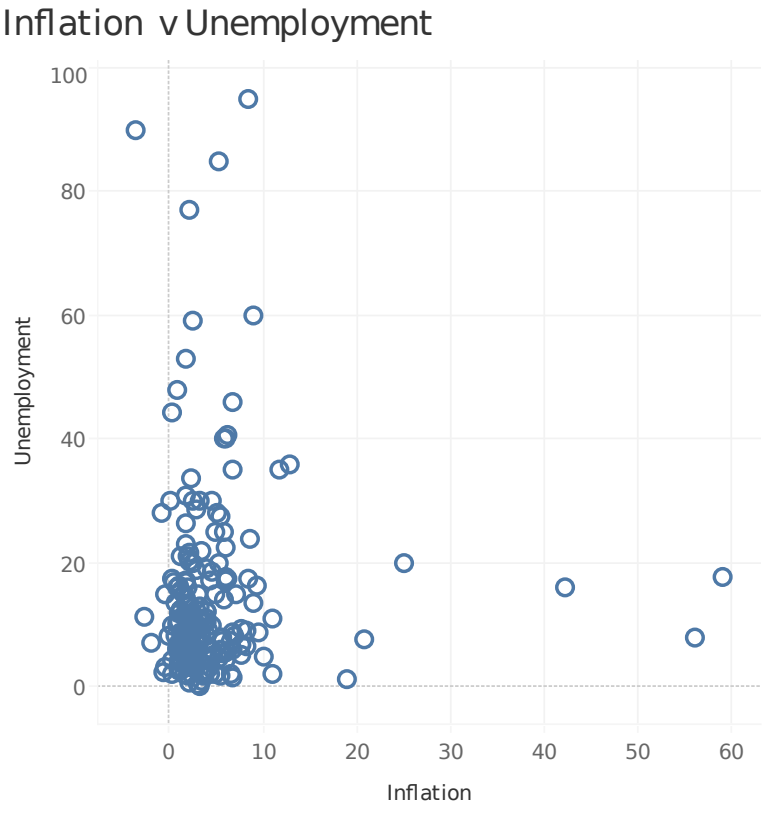
Application to Schema (1)

Mandatory economy(code, inflation, unemployment)



Transformations: Attribute Specialisation

Formalises dropping of null attributes performed in Basic Entity example earlier.

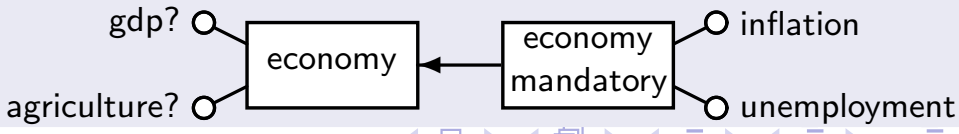


Transformation

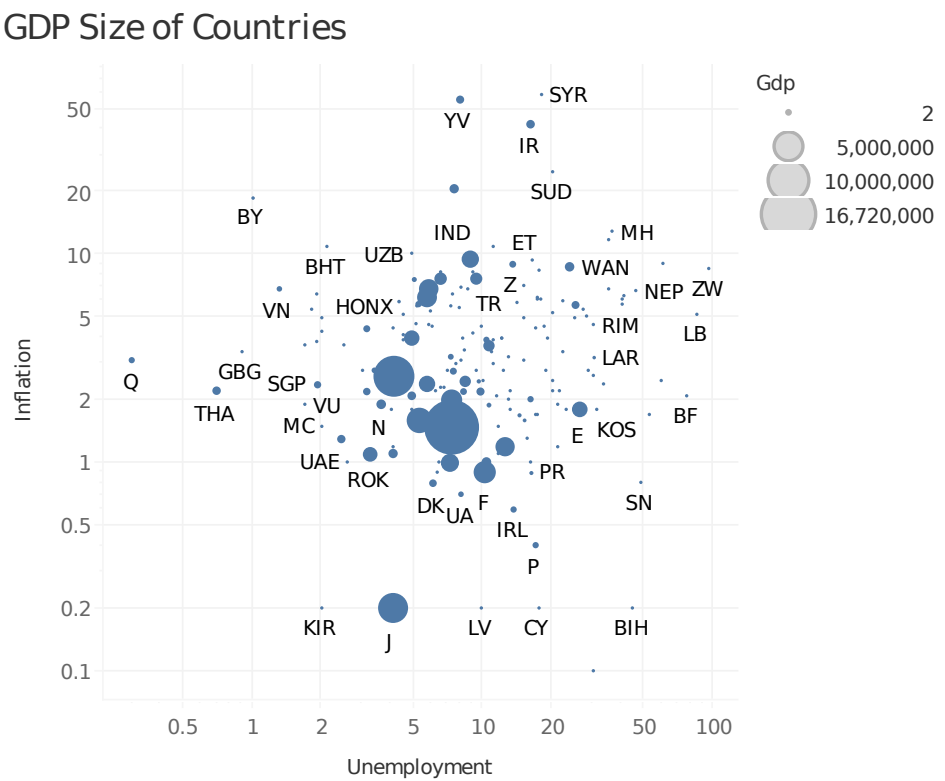
Mandatory $E(k, a_1, \ldots, a_n) \rightarrow$
$EM(k, a_1, \ldots, a_n)$:-
$E(k, a_1, \ldots, a_n)$,
$\text{IsNotNull}(a_1)$,
$\ldots$,
$\text{IsNotNull}(a_n)$.
Not information preserving

Application to Schema (1)

Mandatory economy(code, inflation, unemployment)



Transformations: Attribute Specialisation



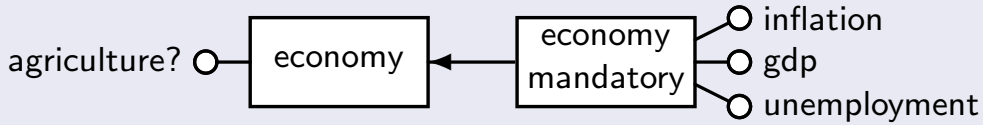
Transformation

Mandatory $E(k, a_1, \ldots, a_n) \rightarrow$
$EM(k, a_1, \ldots, a_n)$:-
$E(k, a_1, \ldots, a_n)$,
$\text{IsNotNull}(a_1)$,
$\ldots$,
$\text{IsNotNull}(a_n)$.

Not information preserving

Application to Schema (2)

Mandatory economy(code, inflation, unemployment, gdp)



Conclusions

We propose a conceptual modelling approach to matching data and visualisations

Conclusions

We propose a conceptual modelling approach to matching data and visualisations

We have identified a small set of visualisation schema patterns that characterise each distinct — from a data representation capability — group of commonly used visualisations

Conclusions

We propose a conceptual modelling approach to matching data and visualisations

We have identified a small set of visualisation schema patterns that characterise each distinct — from a data representation capability — group of commonly used visualisations

We use these visualisation schema patterns and the conceptual schema of the data to create mappings between the data schema and the visualisation schema pattern

Conclusions

We propose a conceptual modelling approach to matching data and visualisations

We have identified a small set of visualisation schema patterns that characterise each distinct — from a data representation capability — group of commonly used visualisations

We use these visualisation schema patterns and the conceptual schema of the data to create mappings between the data schema and the visualisation schema pattern

Domain experts interact with conceptual models of their data to select meaningful visualisations

Conclusions

We propose a conceptual modelling approach to matching data and visualisations

We have identified a small set of visualisation schema patterns that characterise each distinct — from a data representation capability — group of commonly used visualisations

We use these visualisation schema patterns and the conceptual schema of the data to create mappings between the data schema and the visualisation schema pattern

Domain experts interact with conceptual models of their data to select meaningful visualisations

By applying well-known schema transformations, additional visualisation matchings can be generated

Conclusions

We propose a conceptual modelling approach to matching data and visualisations

We have identified a small set of visualisation schema patterns that characterise each distinct — from a data representation capability — group of commonly used visualisations

We use these visualisation schema patterns and the conceptual schema of the data to create mappings between the data schema and the visualisation schema pattern

Domain experts interact with conceptual models of their data to select meaningful visualisations

By applying well-known schema transformations, additional visualisation matchings can be generated

Ongoing work includes implementation of the approach, to be followed by user evaluation