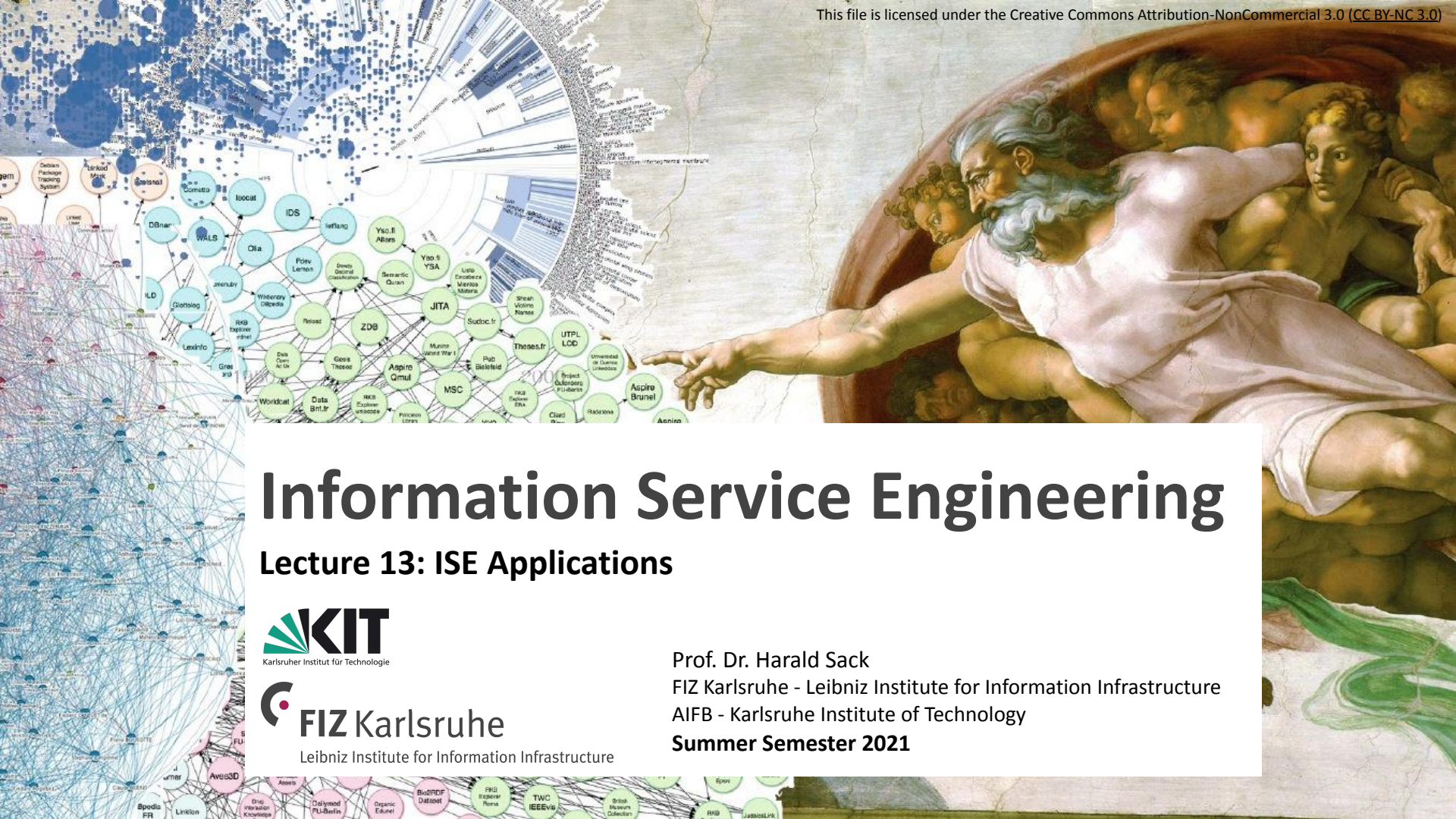


Information Service Engineering

Lecture 13: ISE Applications

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

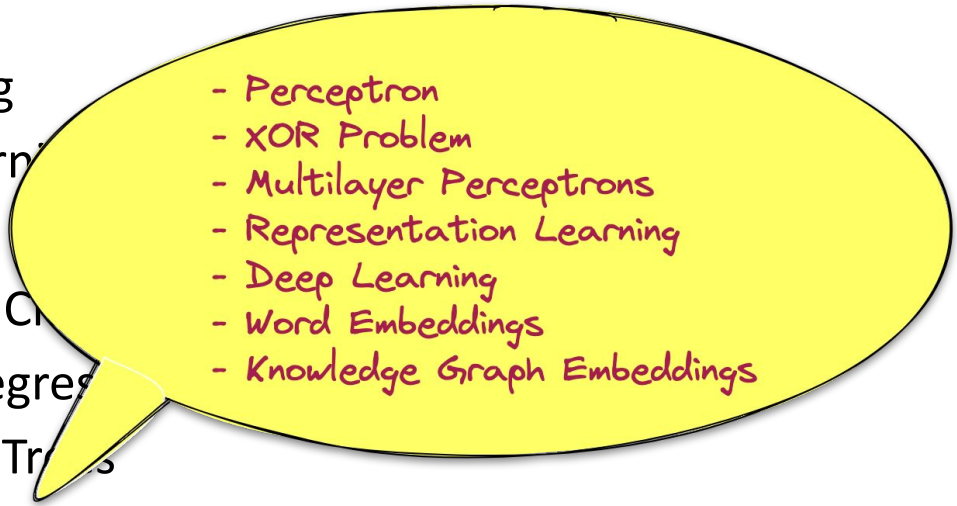
Prof. Dr. Harald Sack
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology

Summer Semester 2021

Information Service Engineering

Last Lecture: Machine Learning - 3

4.1 A Brief History of AI
4.2 Introduction to Machine Learning
4.3 Main Challenges of Machine Learn
4.4 Machine Learning Workflow
4.5 Basic ML Algorithms 1 - k-Means C
4.6 Basic ML Algorithms 2 - Linear Regres
4.7 Basic ML Algorithms 3 - Decision Tr
4.8 Neural Networks and Deep Learning
4.9 Word Embeddings
4.10 Knowledge Graph Embeddings



Information Service Engineering

Lecture Overview

1. Information, Natural Language and the Web
2. Natural Language Processing
3. Knowledge Graphs
4. Basic Machine Learning
5. **ISE Applications**

Information Service Engineering

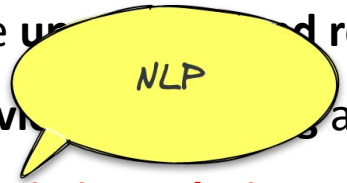
5. ISE Applications

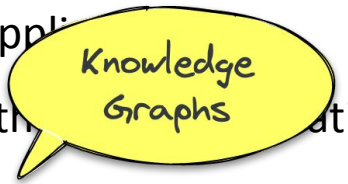
What is Information Service Engineering?

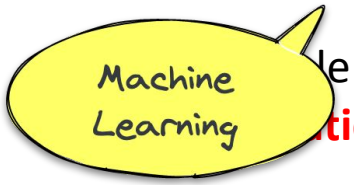
- **Information Service Engineering** investigates **models and methods**
  - to **analyze and integrate structured and unstructured distributed data** from **heterogeneous data sources**
  - with the goal to provide **up-to-date and reliable information services**.
- To this end, **Information Service Engineering** applies
  - both **statistical and linguistic analysis** methods in combination with
  - **machine learning** and **symbolic knowledge representation**
  - to enable the **implementation and sustained provision** of **intelligent information services.**

What is Information Service Engineering?

- **Information Service Engineering** investigates **models and methods**
  - to **analyze and integrate structured and unstructured distributed data** from **heterogeneous data sources**
  - with the goal to provide **u d reliable information services**.
- To this end, **Information Servi** appli
  - both **statistical and linguistic analysis** meth ation with
  - **machine learning** and **knowledge representation**
  - le the **implementation and sustained provision** of **intelligent tion services**.







Information Service Engineering

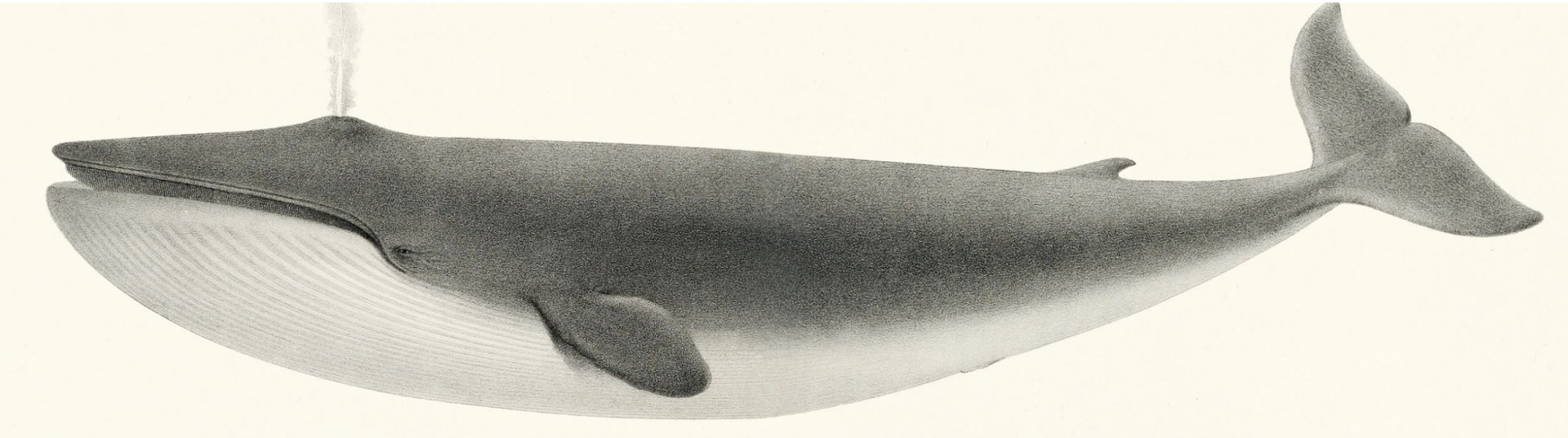
5. ISE Applications

33.6

33.6 m

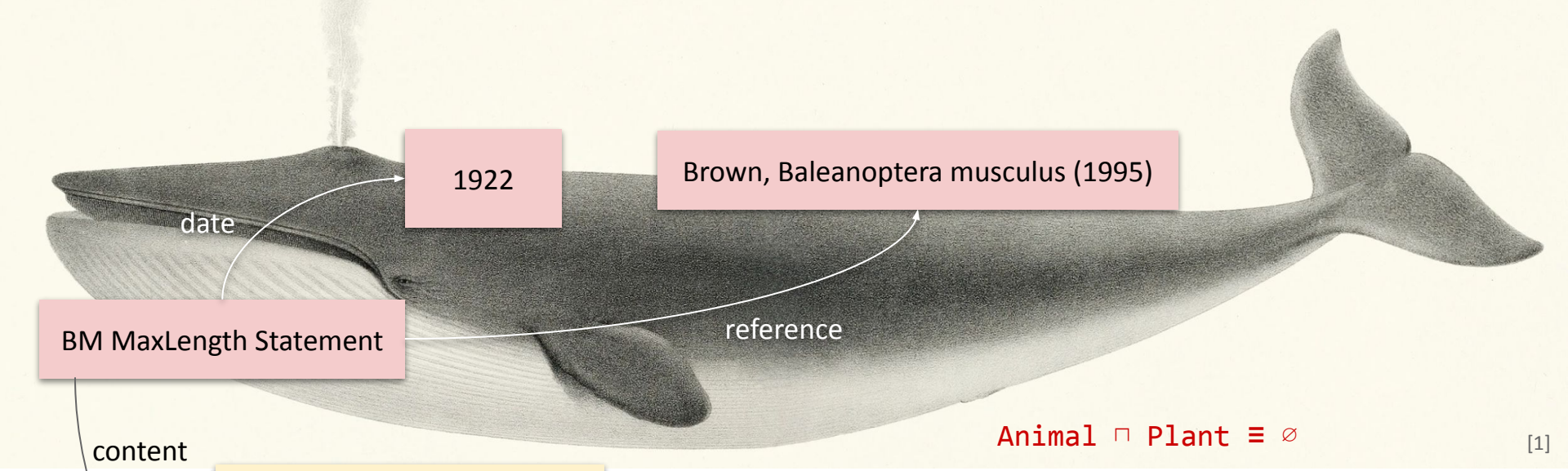

33.6m

33.6m [1] (1922)



[1] S. G. Brown: *Balaenoptera musculus (Linnaeus 1758) – Blauwal*, in Jochen Niethammer, Franz Krapp (Hrsg.): Handbuch der Säugetiere Europas. Band 6: Meeressäuger, Teil I Wale und Delphine – Cetacea, Teil IB: Ziphidae, Kogiidae, Physeteridae, Balaenidae, Balaenopteridae. Aula-Verlag Wiesbaden 1995

Natural history of the cetaceans and other marine mammals of the western coast of North America (1872) by Charles Melville Scammon (1825-1911). [Public Domain]


1922
Brown, Baleanoptera musculus (1995)
date
BM MaxLength Statement
reference
content
Baleanoptera musculus
maxLengthM
33.6
subClass Of
Animal
Plant
subClass Of
subClass Of
Thing
Animal ⊑ Thing
Plant ⊑ Thing
Animal ⊓ Plant ≡ ∅
[1]
BaleanopteraMusculus ⊑ Animal ⊓ ∀maxLengthM.≤33.6

Data

- **Data** is raw.
- It simply exists and has **no significance** beyond its existence (in and of itself).
- It can exist **in any form**, usable or not.

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Information

- **Information** is data that has been given **meaning** by way of **relational connection**.
- This "meaning" can be **useful**, but does not have to be.
- **Information** is **contained in descriptions**, answers to questions that begin with such words as *who, what, when, where, and how many.*

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Knowledge

- **Knowledge** is the appropriate collection of information, such that its intent is to be **useful.**

Wisdom

- **Wisdom is the ability to make sound judgments and decisions**.
- **Data** transforms to **information** by *convention*, **information** to **knowledge** by *cognition*, and **knowledge** to **wisdom** by *contemplation.*

Knowledge Mining

understanding principles
(why?, what is best?, doing things right)

collective application of knowledge in context

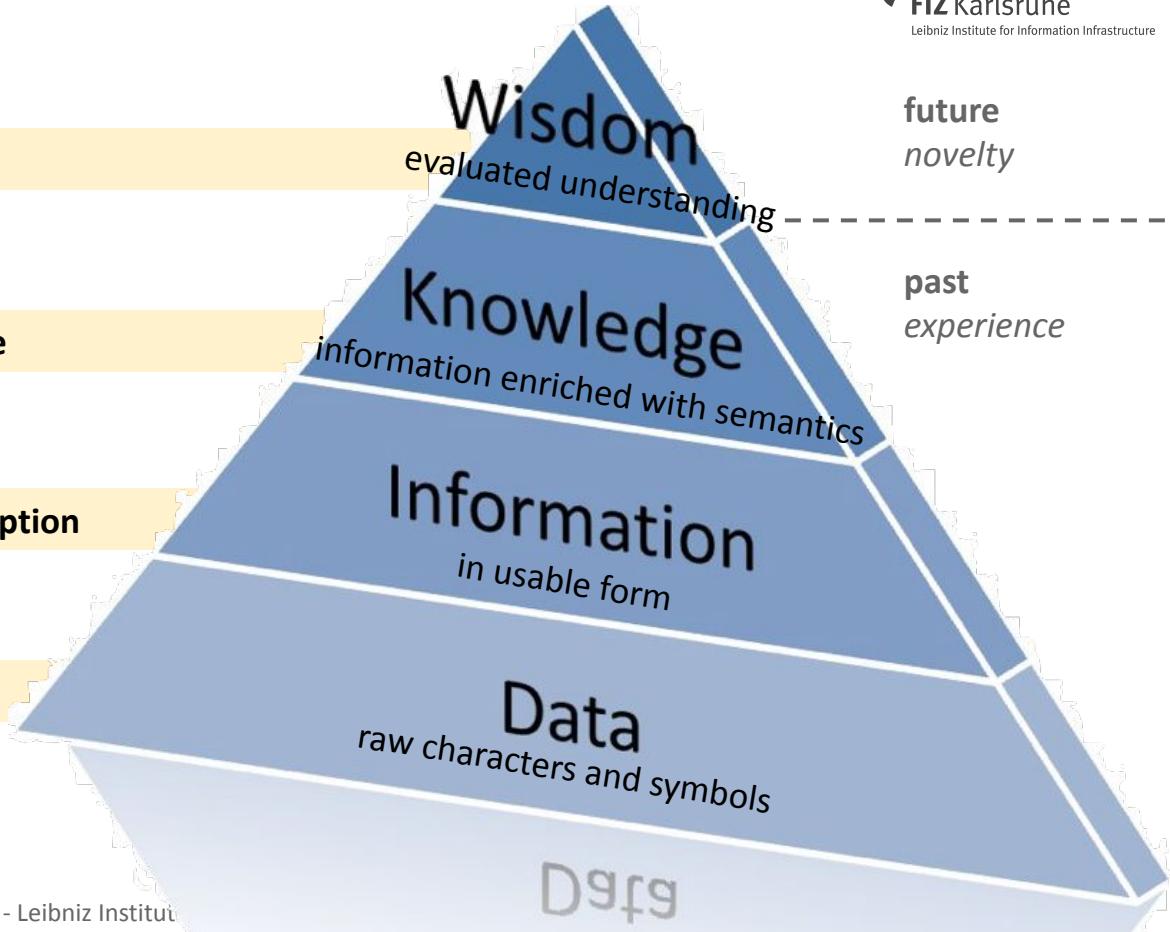
understanding patterns
(principles: how to?)

experience, context, value applied to a message

understanding relations
(description: what?)

a message meant to change the receivers perception

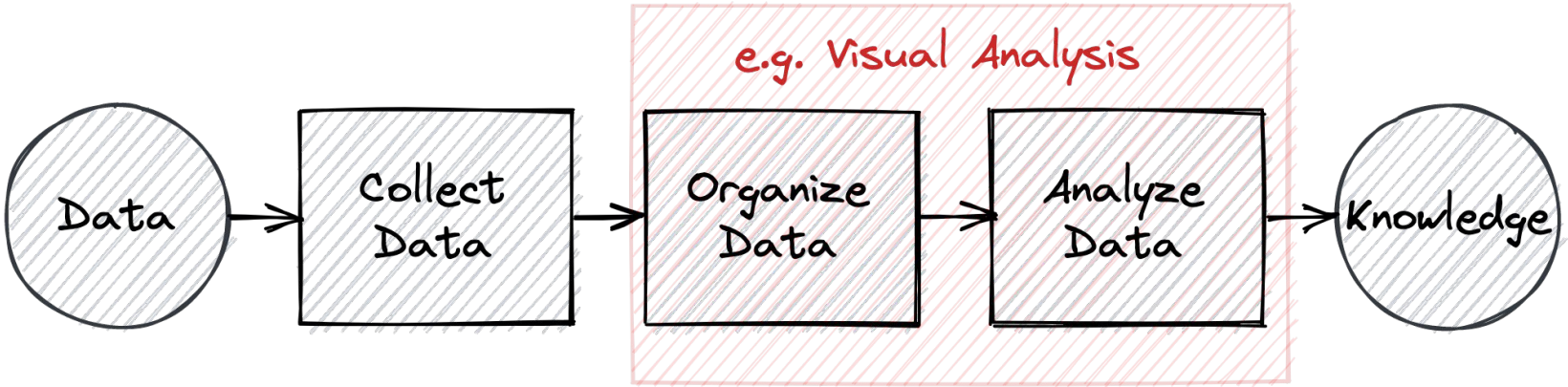
discrete objective facts about event

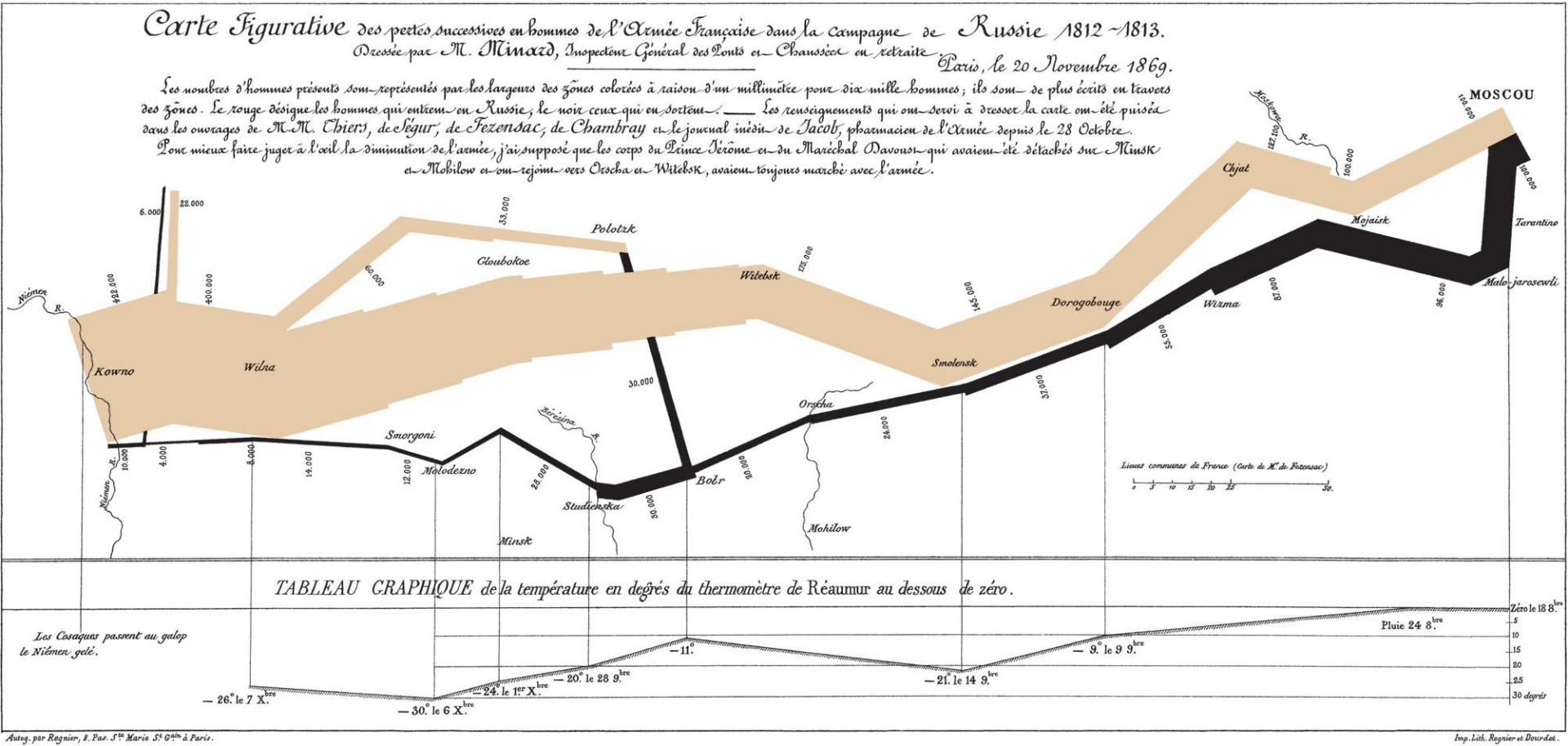


DIKW Pyramid, *Ackoff 1989 [1]*

Data and Knowledge Mining

- How do we transform **data** into **knowledge**?
  1. Collect Data
  2. Organize Data
  3. Analyze Data





Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign,[public domain]

A Picture is Worth a Thousand Words...

- Pictures have been used to convey information long before the development of writing.
- A single picture can be processed (“understood”) much faster than a (linear) text page.
- Human perception is processing in **parallel**, text analysis is limited by the **sequential** process of reading.

Information Visualization

- **Information Visualization** is the study of (interactive) visual representations of abstract data to reinforce human cognition.
- **Information graphics** or **infographics** are graphic visual representations of information, data or knowledge intended to present information quickly and clearly.
- Infographics are a static form of information visualization that aims to emphasize specific findings gained from the visualized data.
- **Mandatory precondition**: Data Analysis.

A Quick Data Visualization Example

Workflow

- **Dataset Generation:**
  Knowledge Graph Mining

- **Task:**
  Draw a **map chart** which visualizes the **number of women soccer players per country**.

e.g. via GoogleSheets or Wikidata

A Quick Data Visualization Example

Dataset Generation

- How to create a SPARQL Query to extract all data about female soccer players from DBpedia?

  1. Look up a famous example you know at Wikipedia

SPARQL Query

extract raw data

Marta (footballer) https://en.wikipedia.org/wiki/Marta_(footballer)

From Wikipedia, the free encyclopedia

This name uses Portuguese naming customs: the first or maternal family name is Vieira *and the second or paternal family name is* da Silva.

Marta Vieira da Silva (born 19 February 1986), commonly known as **Marta** ([ˈmaʁtɐ]), is a Brazilian footballer with both Brazilian and Swedish citizenship.[2] She plays for the Orlando Pride in the National Women's Soccer League and the Brazil national team as a forward. She holds the record for most goals in Brazilian International Football, male or female, with 109 goals for her country. With 17 goals, she also holds the record for most goals scored in the FIFA World Cup tournament (women's or men's).[3][4] Moreover, she is the first footballer of either gender to score at five World Cup editions,[5] a feat matched by Christine Sinclair in 2019.[6] At a club level, Marta won the UEFA Women's Cup at Swedish club Umeå IK in 2004 and won seven Swedish league championships during her time playing for various teams in the country.

Marta is often regarded as the greatest female footballer of all time.[7][8][9][10] She has been named FIFA World Player of the Year six times, five of them being consecutive (from 2006 through 2010) and the latest award coming in 2018. She was a member of the Brazilian national teams that won the silver medal at the 2004 and 2008 Summer Olympics. She was also awarded the Golden Ball (MVP) at the 2004 FIFA U-19 Women's World Championship and won both the Golden Ball award as the best player and the Golden Boot award as the top scorer in the 2007 Women's World Cup after leading Brazil to the final of the tournament.

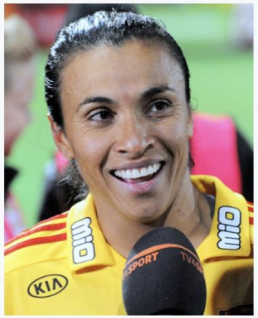
In January 2013 she was named as one of the six Ambassadors of the 2014 FIFA World Cup in Brazil, alongside Amarildo, Bebeto, Carlos Alberto Torres, Ronaldo and Mario Zagallo.[11] She also appeared in the Sveriges Television television documentary series *The Other Sport* from 2013.

In August 2016, Marta was one of the eight to carry the Olympic Flag in the Olympic Games in Rio de Janeiro.

She was appointed by the Secretary-General of the United Nations as a Sustainable Development Goals advocate. The SDG are 17 global goals set with hopes of making the world a better place, and 17 advocates were appointed to help accomplish it.

Contents [hide]

Marta Vieira da Silva

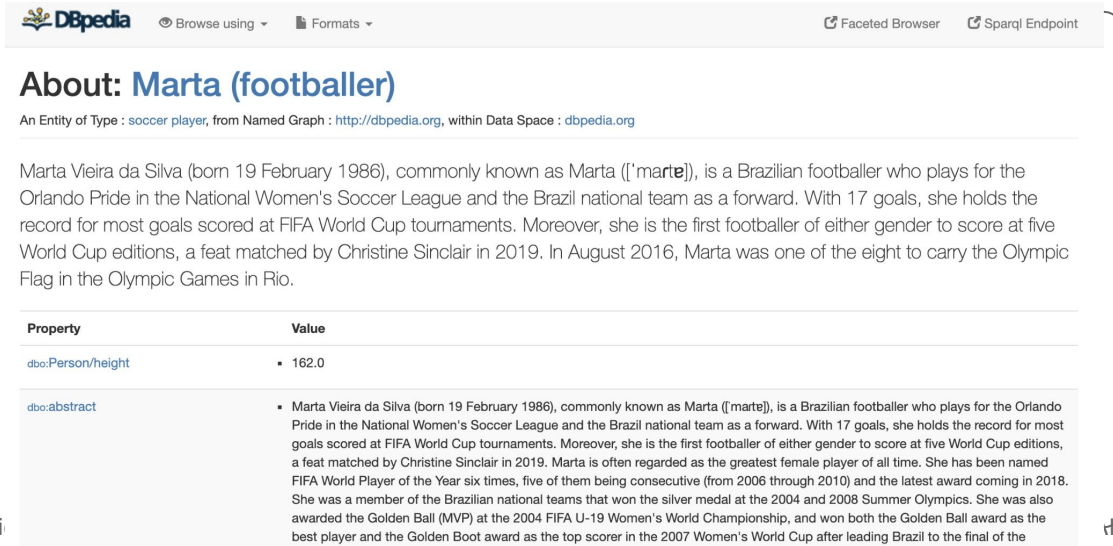


Personal information	
Full name	Marta Vieira da Silva
Date of birth	19 February 1986 (age 35)
Place of birth	Dois Riachos, Alagoas, Brazil
Height	1.62 m (5 ft 4 in)[1]
Position(s)	Forward

A Quick Data Visualization Example

Dataset Generation

- How to create a SPARQL Query to extract all data about female soccer players from DBpedia?
  1. Look up a famous example you know at Wikipedia.
  2. Look up the same at DBpedia. https://dbpedia.org/page/Marta_(footballer)



SPARQL Query

extract raw data

check for available properties to extract the desired data

This is not always straight forward

A Quick Data Visualization Example

Dataset Generation

- **Compose the SPARQL Query:**
  - **Country** and **Number of woman soccer players per country**

woman soccer player

```
?s dct:subject/skos:broader*
   dbc:Women\'s_association_football_players .
```

country

```
?s dbo:birthPlace ?birthplace .
?birthplace dbo:country ?country .
?country rdf:type dbo:Country .
?country rdfs:label ?countryLabel
FILTER (lang(?countryLabel)="en")
```

group and count by country

```
GROUP BY ?countryLabel
COUNT(DISTINCT ?s)
```

SPARQL Query Editor About Tables Conductor Facet Browser Permalink

Extensions: cxml save to dav sponge User: SPARQL

Default Data Set Name (Graph IRI)

http://dbpedia.org

Query Text

SPARQL query

```
select ?clabel (COUNT (DISTINCT ?s) as ?count) WHERE {
?s dct:subject/skos:broader* dbc:Women\'s_association_football_players ;
  dbo:birthPlace ?birthplace .
?birthplace dbo:country ?country .
?country rdfs:label ?countryLabel FILTER (lang(?countryLabel)="en").
BIND (STR(?countryLabel) as ?clabel)
}
GROUP BY ?clabel
ORDER BY DESC(?count)
```

Results Format HTML

Execute Query Reset

Execution timeout 30000 milliseconds

A Quick Data Visualization Example

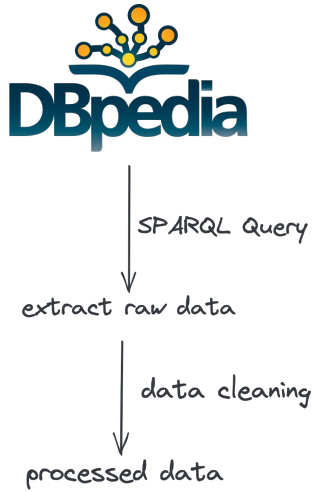
Dataset Generation

SPARQL | HTML5 table

clabel	count
United States	513
United Kingdom	372
Australia	188
Spain	171
Japan	159
France	148
Turkey	129
Sweden	116
Brazil	103
Mexico	101
Netherlands	84
Italy	83
Canada	59
Denmark	55
Czech Republic	50
Argentina	46
Republic of Ireland	44

Further **Data Cleaning** required:

- Remove/aggregate countries that are no countries
- Remove/aggregate countries that don't exist anymore or have been replaced by another country
- and probably more...

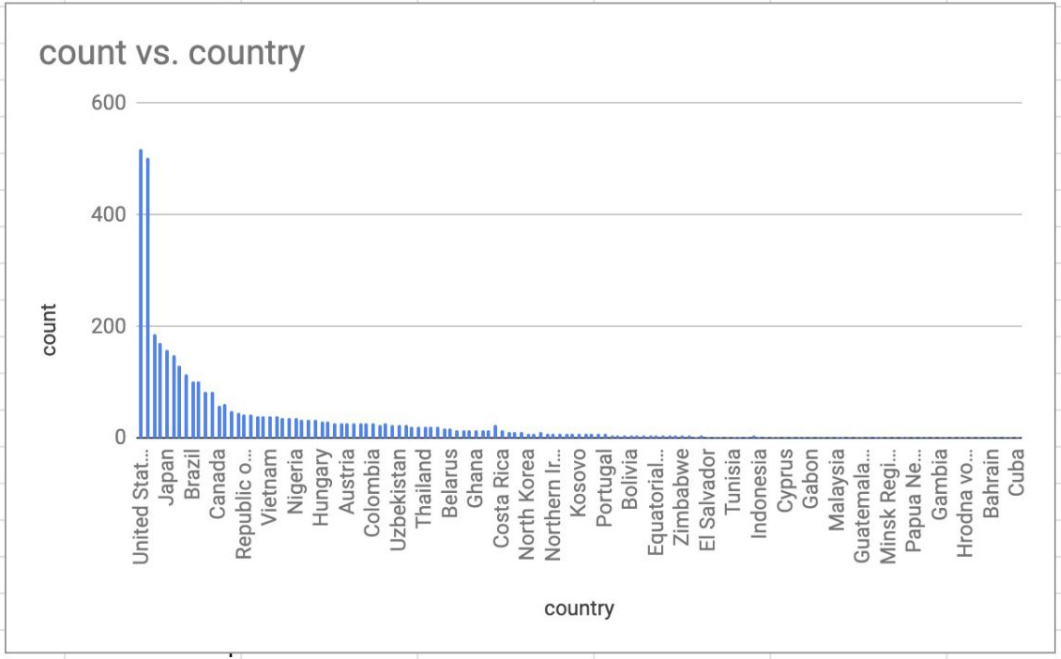


13 - ISE2020 - Soccer Players per Country

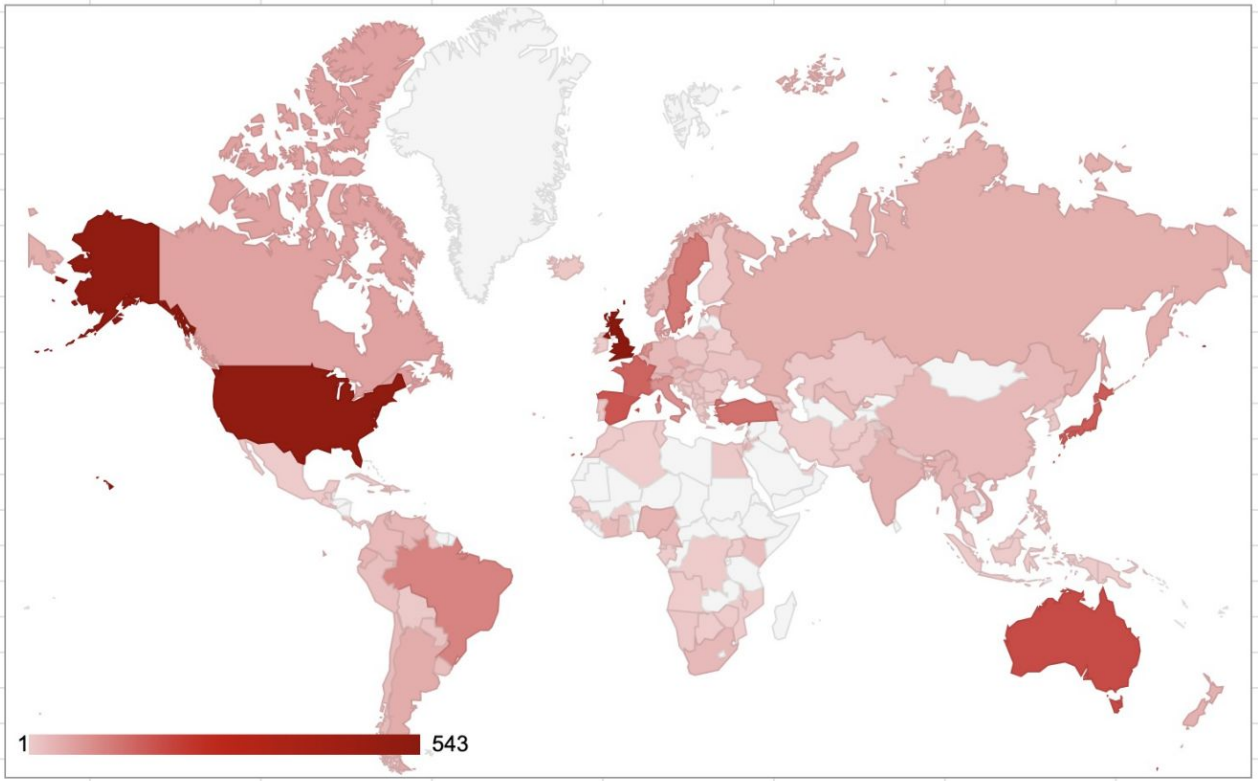
File Edit View Insert Format Data Tools Add-ons Help Last edit was seconds ago

GoogleDoc Sheet with Data

	A	B
64	Bosnia and Herzegovina	10
65	Kazakhstan	8
66	Northern Ireland	8
67	Croatia	8
68	Namibia	
69	Bangladesh	
70	Kosovo	
71	Montenegro	
72	Taiwan	
73	Pakistan	
74	Portugal	
75	Romania	
76	Nepal	
77	South Korea	
78	Bolivia	
79	Moldova	
80	Israel	
81	Egypt	
82	Equatorial Guinea	
83	Dominican Republic	
84	Greece	
85	Afghanistan	
86	Zimbabwe	4
87	Bulgaria	4



North Korea	9
Iceland	9
Kazakhstan	8
Croatia	8
Namibia	8
Bangladesh	8
Kosovo	7
Montenegro	7
Taiwan	7
Pakistan	7
Portugal	7
Romania	6
Nepal	6
South Korea	6
Bolivia	6
Moldova	6
Israel	6
Egypt	6
Equatorial Guinea	6
Dominican Republic	6
Greece	5
Georgia	5
Afghanistan	4
Zimbabwe	4
Bulgaria	4
Congo	4
Lithuania	3



Wikidata Query Service | Examples | Help | More tools | English

SPARQL query

```
#defaultView:Map
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbc: <http://dbpedia.org/resource/Category:>
PREFIX dbo: <http://dbpedia.org/ontology/>

SELECT  ?wditem ?wditemLabel ?image ?birthplaceLabel ?coord WHERE {
  ?wditem wdt:P106 wd:Q937857; # occupation  association football player
          wdt:P21 wd:Q6581072 ; # gender. female
          wdt:P19 ?birthplace . # birthplace
  ?birthplace wdt:P625 ?coord . # located at
  ?wditem wdt:P18 ?image .      # image
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

Table | 3225 results in 54674 ms | Code | Download | Link

Search

wditem	wditemLabel	image	birthplaceLabel	coord
wd:Q252588	Ásta Árnadóttir	commons:Ásta Árnadóttir.jpg	Iceland	Point(-19.0 65.0)
wd:Q273356	Britta Carlson	commons:20150426 PSG vs Wolfsburg 009.jpg	Kiel	Point(10.139444444 54.323333333)
wd:Q242217	Inka Grings	commons:2018-06-24 Inka Grings-9889.jpg	Düsseldorf	Point(6.772380555 51.231144444)

SPARQL query

Dois Riachos
Point(-37.100833333 -9.392777777)

Marta

Information Service Engineering

5. ISE Applications

Knowledge Mining and Knowledge Discovery

Definitions

Knowledge Discovery [in Databases] (KDD) is the nontrivial process of identifying **valid**, **novel**, **potentially useful**, and **ultimately understandable patterns** in (massive) data sources.
(Fayyad et al, 1996)

- **valid**: to a certain degree the discovered patterns should also hold for new, previously unseen problem instances.
- **novel**: at least to the system and preferable to the user.
- **potentially useful**: they should lead to some benefit to the user or task.
- **ultimately understandable**: the end user should be able to interpret the patterns either immediately or after some post-processing.

Knowledge Mining and Knowledge Discovery

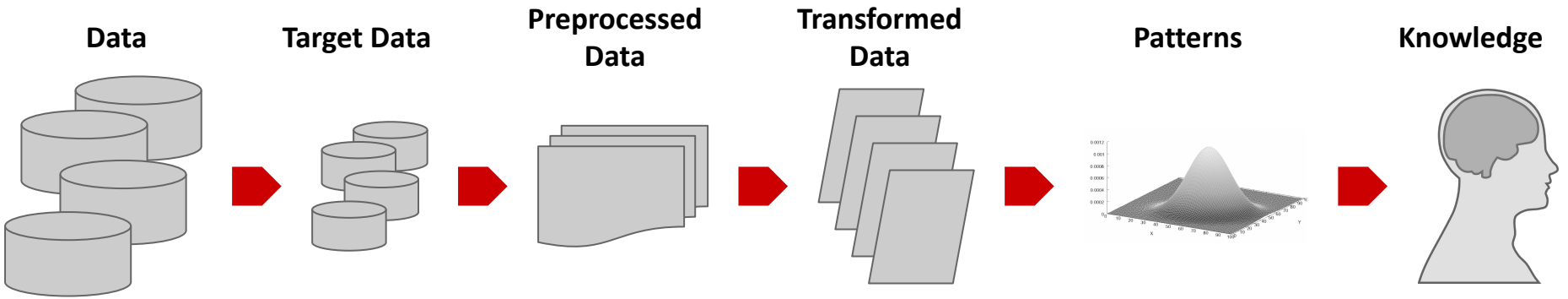
Goals

Knowledge Discovery [in Databases] (KDD) is the nontrivial process of identifying **valid**, **novel**, **potentially useful**, and **ultimately understandable patterns** in (massive) data sources.
(Fayyad et al, 1996)

- **Goals:**
  - **Descriptive Modelling:** explains the characteristic and the behaviour of the observed data.
  - **Predictive Modelling:** predicts the behaviour of new data based on some model.
- **Important:**
  - The extracted model/pattern does not have to apply in 100% of the cases.

Knowledge Mining and Knowledge Discovery

Process Workflow



Selection:
Select a relevant dataset or focus on a subset of a dataset.

Preprocessing/ Cleaning:
Data integration from different sources, Data Cleaning.

Transformation:
Select useful features, feature transformation, dimensionality reduction.

Data Mining:
Search for patterns of interest.

Evaluation:
Evaluate patterns based on interestingness measures, model validation.

Data Cleaning

- **“Dirty” Data:**
  - Dummy values, absence of data, contradicting data, etc.
- **Steps in Data Cleaning**
  - **Parsing**: locates and identifies individual data elements in raw data.
  - **Correcting**: corrects parsed individual data components using sophisticated data algorithms.
  - **Normalization**: applies conversion routines to transform data into standard formats.
  - **Matching**: searching and matching records within and across data based on predefined rules.
  - **Consolidating**: merges data into one representation.

Knowledge Mining Functionality

- **Characterization:** summarizing general features of objects in a target class (concept description).
- **Discrimination:** comparing general features of objects between a target class and a contrasting class (concept comparison).
- **Association:** studying the frequency of items occurring together.
- **Prediction:** predicting some unknown or missing attribute values.
- **Classification:** organizing data in given classes based on attribute values (supervised).
- **Clustering:** organizing data in classes based on attribute values (unsupervised).
- **Outlier analysis:** identifying and explaining exceptions (surprises).
- **Time-series analysis:** analyzing trends and deviations.

Data Analysis

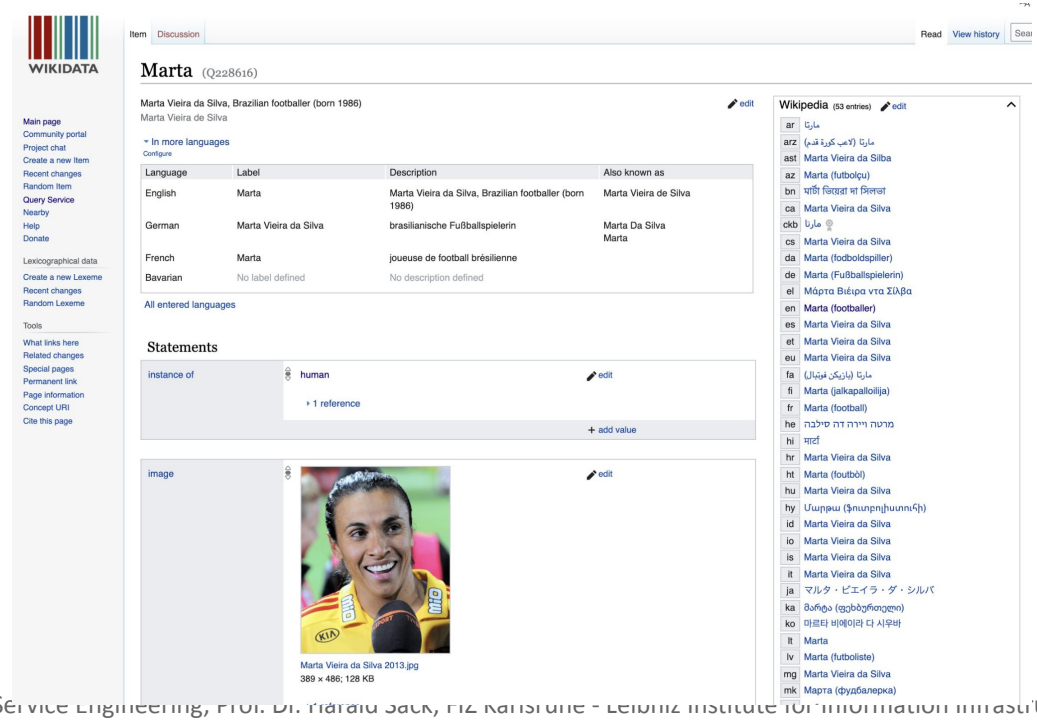
- **Data Analysis** is a fundamental iterative process:
  1. Formulation and execution of a query
  2. Analysis of the results
  3. Formulation of a consecutive query based on the achieved results

- **Goals of Data Analysis:**
  - Maximize understanding of analyzed data
  - Uncover hidden structures/patterns
  - Extraction of important variables
  - Detection of anomalies and outliers
  - Testing of hypotheses
  - Development of a simple model

Let's Analyze (More) Soccer Data

1. **Data Acquisition**

   Look up a sample of the data you want to collect

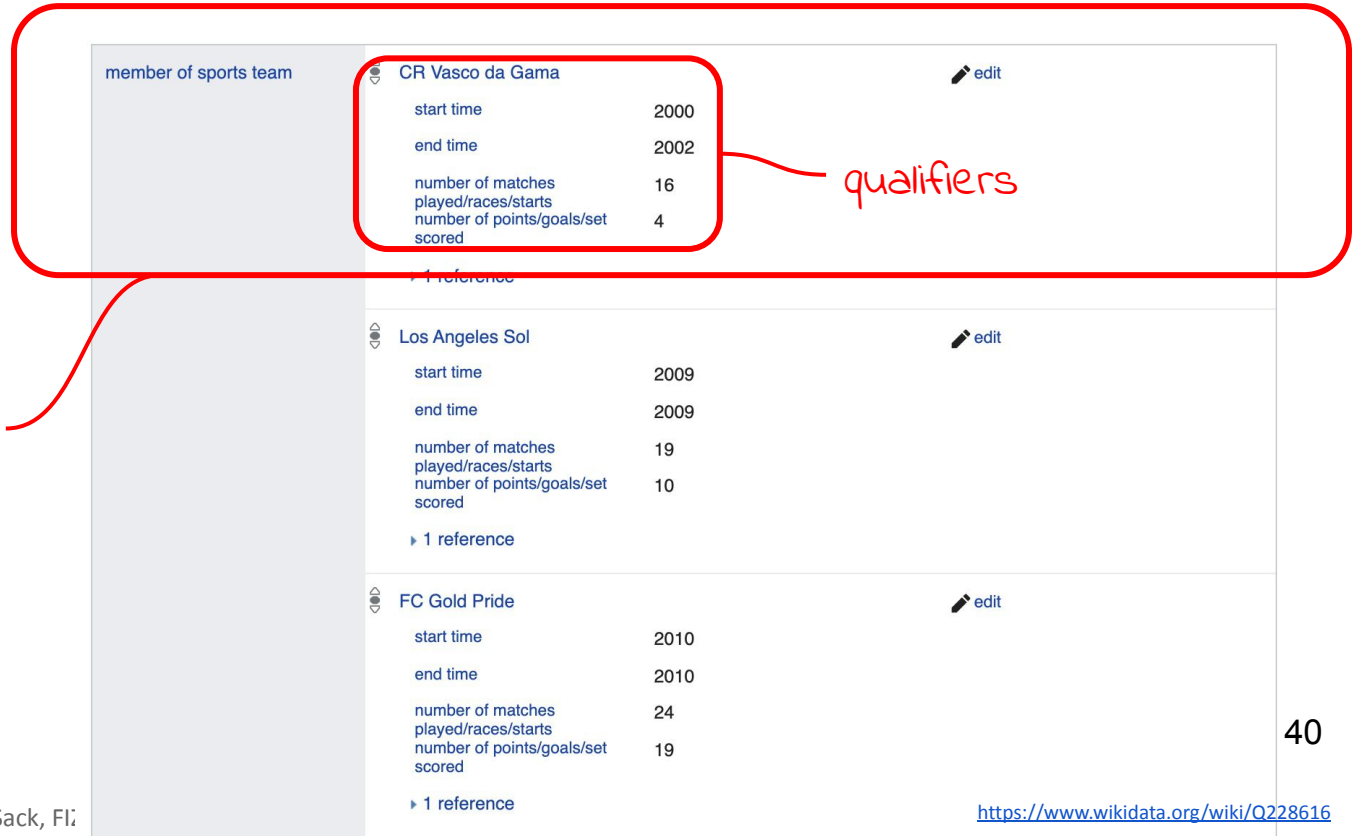


[Wikidata sample page]

Let's Analyze (More) Soccer Data

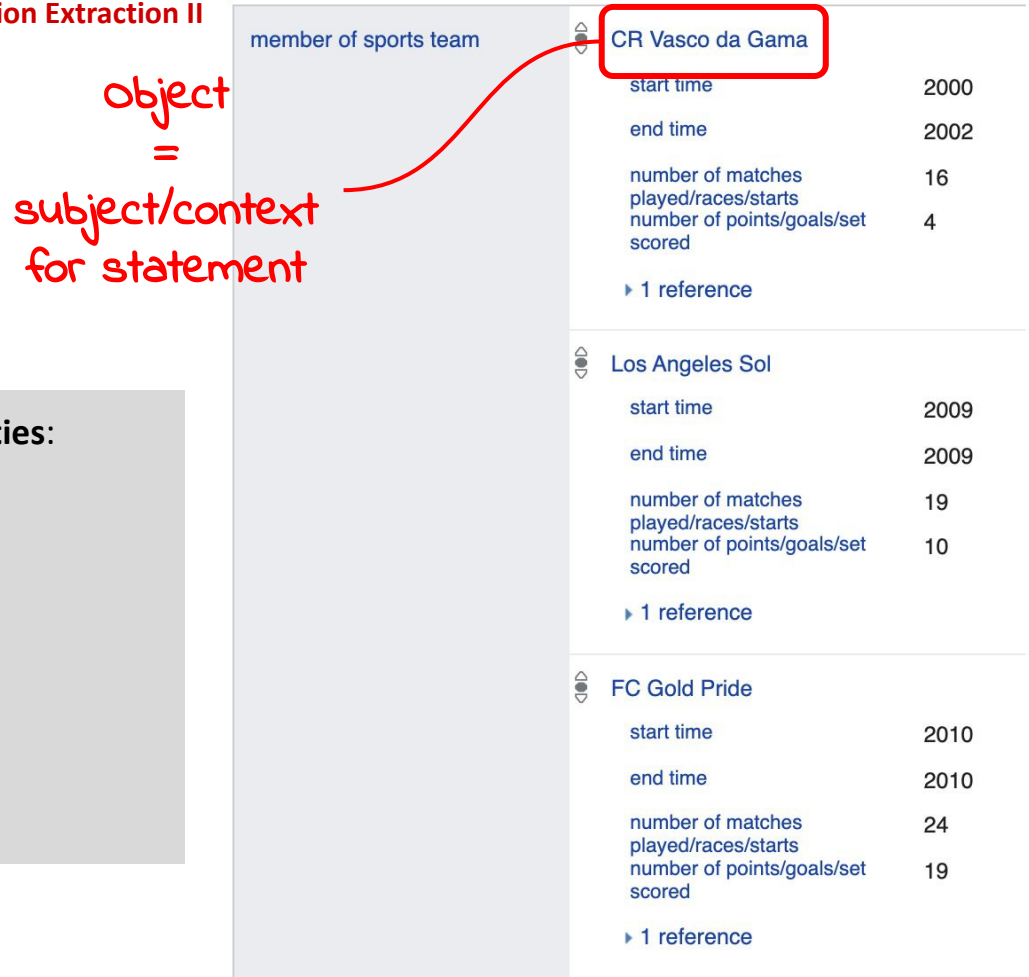
Wikidata Recap

1. **Data Acquisition**
   Look up a sample of the data you want to collect



https://www.wikidata.org/wiki/Q228616

WikiData Recap



object = subject/context for statement

Access via **different namespaces for properties**:

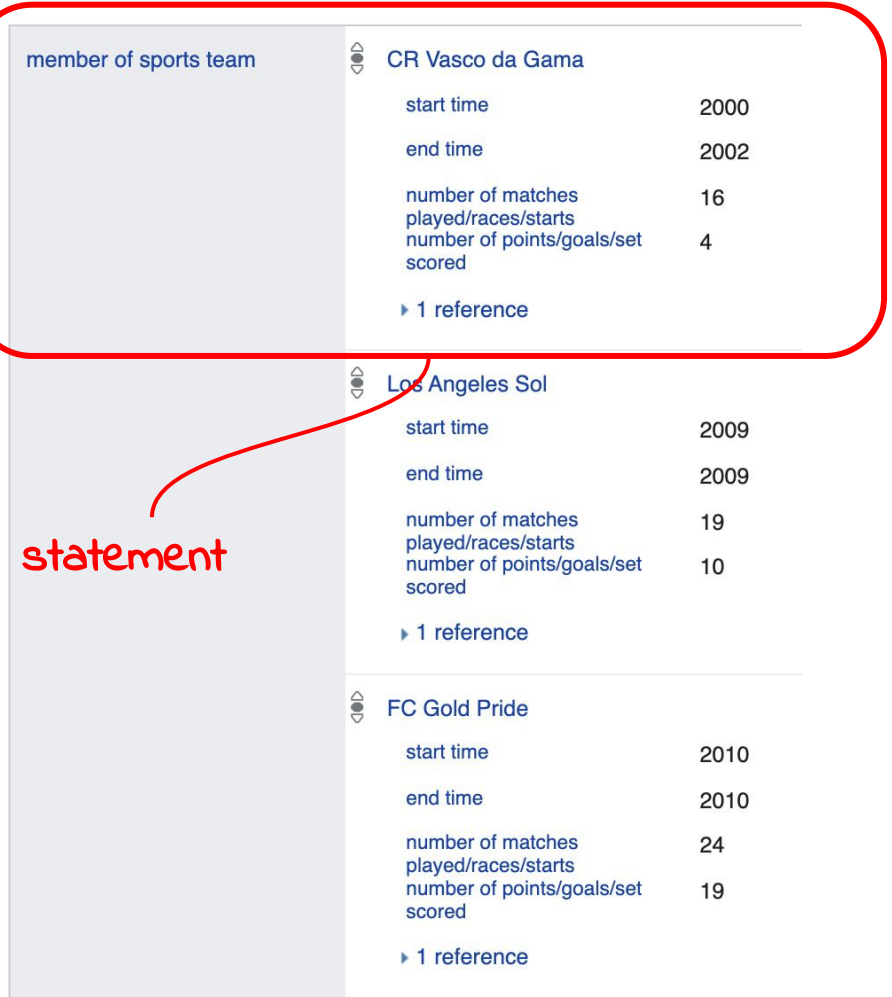
- **wdt:** connects an item to a value
  wd:Q228616 wdt:P54 ?team .

https://www.wikidata.org/wiki/Q228616

WikiData Recap

Access via **different namespaces for properties**:

- **wdt:** connects an item to a value
  wd:Q228616 wdt:P54 ?team .
- **p:** connects a subject to a statement
  wd:Q228616 p:P54 ?team_statement .



https://www.wikidata.org/wiki/Q228616

WikiData Recap

Access via **different namespaces for properties**:

- **wdt:** connects an item to a value
  wd:Q228616 wdt:P54 ?team .
- **p:** connects a subject to a statement
  wd:Q228616 p:P54 ?team_statement .
- **pq:** connects statement to qualifier value
  ?team_statement pq:1351 ?statement_value

https://www.wikidata.org/wiki/Q228616

Let's Analyze (More) Soccer Data

2. **Get Data**

   Interesting facts about Woman Assiciation Soccer Players to extract from Wikidata:

   - has played in how many different **teams**?
   - has played in how many **matches**?
   - has scored how many **goals**?
   - has additional **occupation(s)**?
   - has additional **citizenships**?
   - has how many **Wikipedia pages** (in different languages)?
   - has played on which **position**?
   - **weight**
   - **height**
   - **birthdate**

Let's create a (complicated) SPARQL query

Let's Analyze (More) Soccer Data

2. Get Data

Compose a Wikidata SPARQL query according to our needs.

```
SELECT (COUNT(?team_statement) as ?teams) (SUM(?goals) as ?total_goals)
       (SUM(?matches) as ?total_matches) (COUNT( DISTINCT ?citizenship) as ?total_citizenships)
       (SAMPLE(?height) as ?height)  (SAMPLE(?weight) as ?weight) (SAMPLE(?posLabel) as ?pos)
       (SAMPLE(xsd:date(?birthdate)) as ?bday) (COUNT(DISTINCT ?occupation) as ?sidejobs)
       (SUM(?link) as ?importance)
WHERE {
  ?s wdt:P106 wd:Q937857 ;
     wdt:P21 wd:Q6581072 ;
     p:P54 ?team_statement ;
     wdt:P2048 ?height ;
     wdt:P2067 ?weight ;
     wdt:P413 ?pos ;
     wdt:P569 ?birthdate ;
     wdt:P27 ?citizenship ;
     wdt:P106 ?occupation ;
     wikibase:sitelinks ?link.
  ?team_statement pq:P1351 ?goals ;
                  pq:P1350 ?matches .
  ?pos rdfs:label ?posLabel FILTER (lang(?posLabel)="en") .
} GROUP BY ?s
```

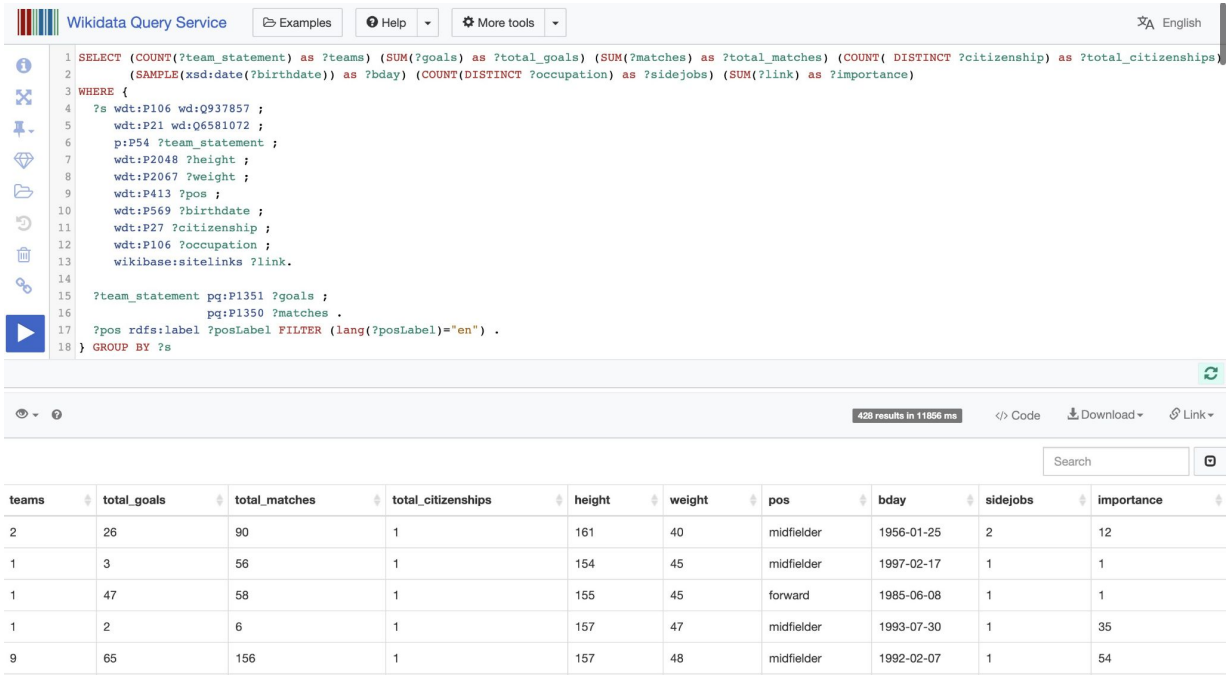
SPARQL query

Let's Analyze (More) Soccer Data

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

2. Get Data

Compose a Wikidata SPARQL query according to our needs.



teams	total_goals	total_matches	total_citizenships	height	weight	pos	bday	sidejobs	importance
2	26	90	1	161	40	midfielder	1956-01-25	2	12
1	3	56	1	154	45	midfielder	1997-02-17	1	1
1	47	58	1	155	45	forward	1985-06-08	1	1
1	2	6	1	157	47	midfielder	1993-07-30	1	35
9	65	156	1	157	48	midfielder	1992-02-07	1	54

SPARQL query

Let's Analyze (More) Soccer Data

2. **Get Data**
   Saving Data...

[Data in Google Doc Spreadsheet]

13 - ISE2021 - Soccer Stats Example

File Edit View Insert Format Data Tools Add-ons Help Last edit was 2 hours ago

	C	D	E	F	G	H	I	J
1	total_matches	otal_citizenship	height	weight	pos	bday	sidejobs	importance
2	90	1	161	40	midfielder	1956-01-25	2	12
3	56	1	154	45	midfielder	1997-02-17	1	1
4	58	1	155	45	forward	1985-06-08	1	1
5	6	1	157	47	midfielder	1993-07-30	1	35
6	156	1	157	48	midfielder	1992-02-07	1	54
7	128	1	157	48	midfielder	1996-03-10	1	2
8	141	1	152	48	midfielder	1979-03-03	1	3
9	3	1	158	48	midfielder	1990-09-27	1	37
10	351	1	160	48	midfielder	1983-09-02	1	18
11	78	1	157	48	midfielder	1993-10-17	1	2
12	179	1	160	47	midfielder	1997-08-18	1	90
13	216	1	1.64	49	midfielder	1984-03-03	2	56
14	233	1	171	50	midfielder	1992-04-18	1	168
15	333	1	154	50	midfielder	1983-03-14	1	25
16	126	1	162	50	midfielder	1977-10-13	2	30
17	17	1	160	50	midfielder	1992-11-04	1	6
18	18	1	166	50	midfielder	1992-05-28	1	1
19	6	1	160	48	midfielder	1998-05-13	1	6
20	164	1	157	52	midfielder	1985-01-28	1	258
21	360	1	162	52	midfielder	1980-06-08	1	84
22	38	2	155	52	midfielder	1990-12-13	1	16
23	194	1	162	51	midfielder	1989-05-14	1	56
24	204	1	157	51	midfielder	1986-04-09	1	28
25	40	1	161	51	midfielder	1994-03-06	1	1
26	114	1	160	50	midfielder	1979-06-11	1	3

Let's Analyze (More) Soccer Data

3. **CleanUp Data**
   This might require
   Several rounds...

[Data in Google Doc Spreadsheet]

13 - ISE2021 - Soccer Stats Example

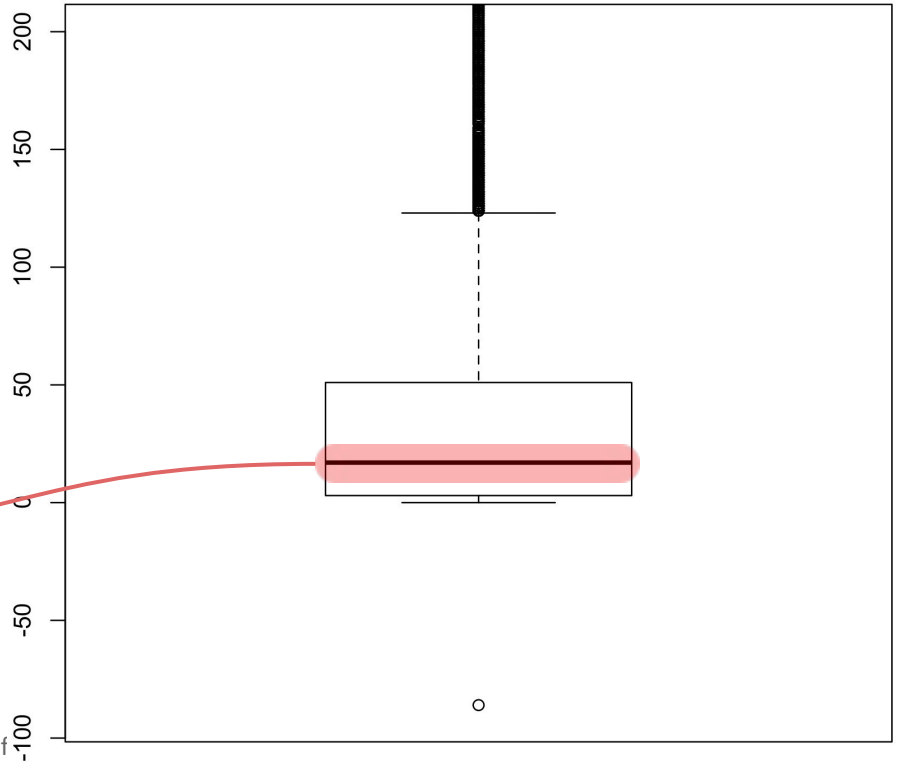
File Edit View Insert Format Data Tools Add-ons Help Last edit was 2 hours ago

	C	D	E	F	G	H	I	J
1	total_matches	otal_citizenship	height	weight	pos	bday	sidejobs	importance
2	90	1	161	40	midfielder	1956-01-25	2	12
3	56	1	154	45	midfielder	1997-02-17	1	1
4	58	1	155	45	forward	1985-06-08	1	1
5	6	1	157	47	midfielder	1993-07-30	1	35
6	156	1	157	48	midfielder	1992-02-07	1	54
7	128	1	157	48	midfielder	1996-03-10	1	2
8	141	1	152	48	midfielder	1979-03-03	1	3
9	3	1	158	48	midfielder	1990-09-27	1	37
10	351	1	160	48	midfielder	1983-09-02	1	18
11	78	1	157	48	midfielder	1993-10-17	1	2
12	179	1	160	47	midfielder	1997-08-18	1	90
13	216	1	1.64	49	midfielder	1984-03-03	2	56
14	233	1	171	50	midfielder	1992-04-18	1	168
15	333	1	154	50	midfielder	1983-03-14	1	25
16	126	1	162	50	midfielder	1977-10-13	2	30
17	17	1	160	50	midfielder	1992-11-04	1	6
18	18	1	166	50	midfielder	1992-05-28	1	1
19	6	1	160	48	midfielder	1998-05-13	1	6
20	164	1	157	52	midfielder	1985-01-28	1	258
21	360	1	162	52	midfielder	1980-06-08	1	84
22	38	2	155	52	midfielder	1990-12-13	1	16
23	194	1	162	51	midfielder	1989-05-14	1	56
24	204	1	157	51	midfielder	1986-04-09	1	28
25	40	1	161	51	midfielder	1994-03-06	1	1
26	114	1	160	50	midfielder	1979-06-11	1	3

Excursion: Learn How to Read Boxplots

The **median** is the value separating the higher half from the lower half of a data sample.

```
       X0
Min.   : -86.00
1st Qu.:   3.00
Median :  17.00
Mean   :  38.85
3rd Qu.:  51.00
Max.   :1329.00
```

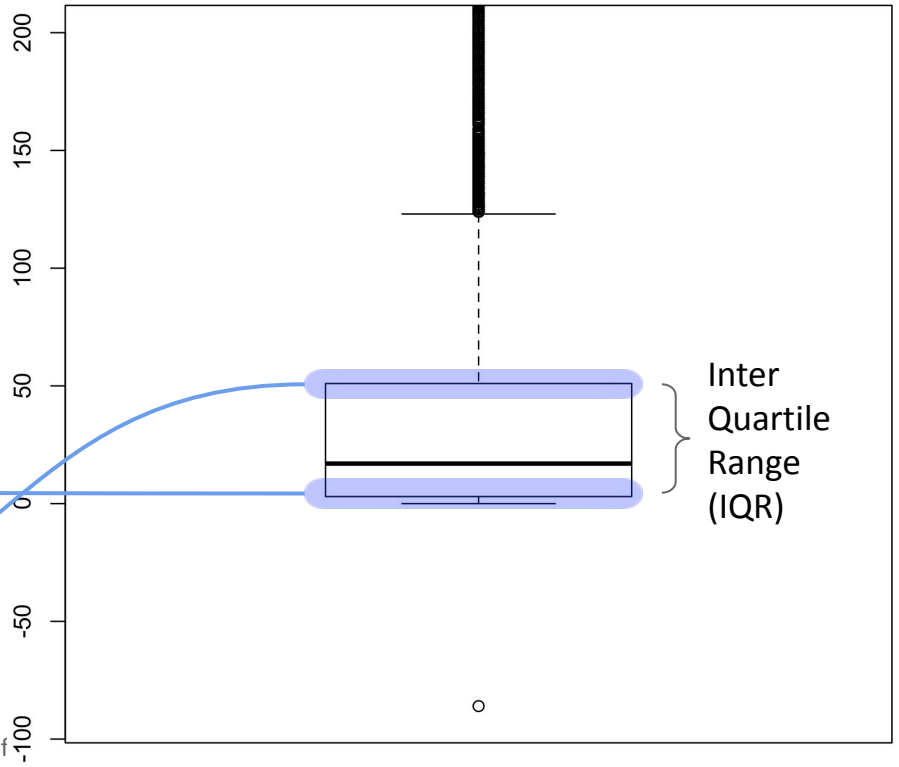


Excursion: Learn How to Read Boxplots

The **first quartile (Q_1)** is defined as the middle number between the smallest number and the median of the data set.

The **third quartile (Q_3)** is the middle value between the median and the highest value of the data set.

```
       X0
Min.   : -86.00
1st Qu.:   3.00
Median :  17.00
Mean   :  38.85
3rd Qu.:  51.00
Max.   :1329.00
```



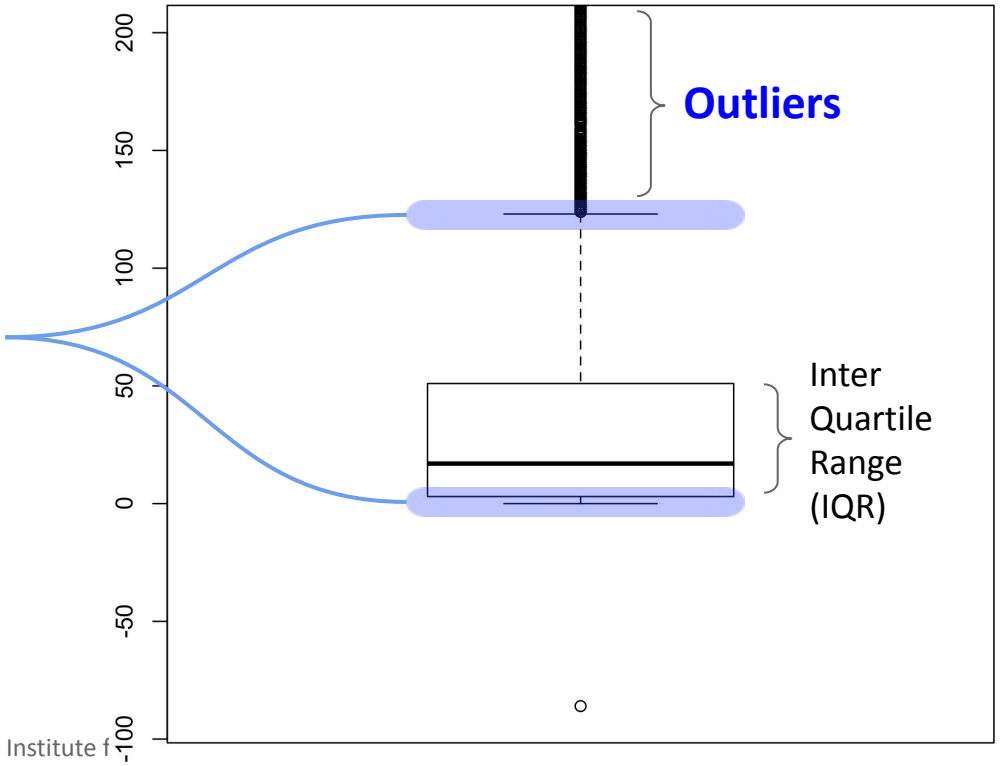
Excursion: Learn How to Read Boxplots

Whiskers are indicating variability outside the upper and lower quartiles.

Any data not included between the whiskers should be considered as an **outlier**.

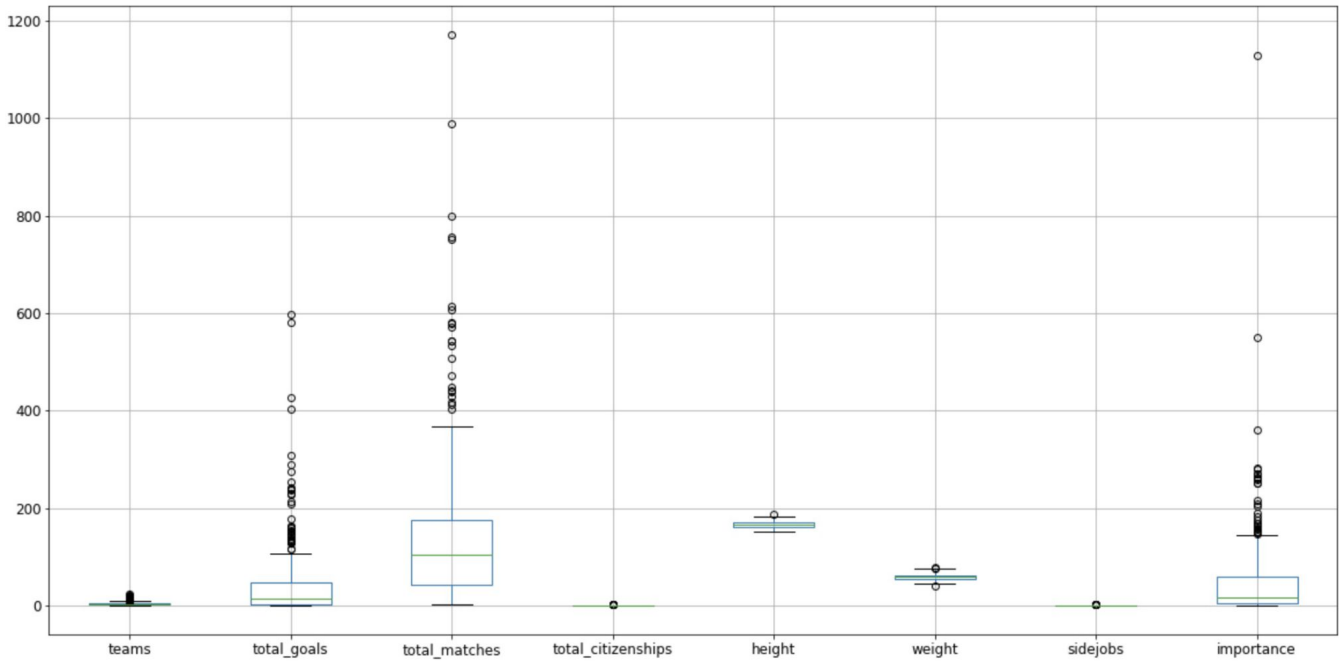
Whiskers: IQR x 1.5 = (Q_3-Q_1) x 1.5

```
       X0
Min.   : -86.00
1st Qu.:   3.00
Median :  17.00
Mean   :  38.85
3rd Qu.:  51.00
Max.   :1329.00
```



Excursion: Learn How to Read Boxplots

4. **Analyse the Data**
   E.g. via python



Data in Google Collab Notebook

Information Service Engineering

5. ISE Applications

Information Service Engineering

5. ISE Applications

5. ISE Applications
Bibliography

- Ackoff, R. L. (1989). ***From data to wisdom***. Journal of Applied Systems Analysis 15: 3-9
- Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. (1996). ***From data mining to knowledge discovery: an overview***. In Advances in knowledge discovery and data mining, American Association for Artificial Intelligence, Menlo Park, CA, USA 1-34.

5. ISE Applications

Syllabus Questions

- What's the difference: Data, Information, Knowledge, and Wisdom?
- How do we get from Data to Information, from Information to Knowledge, and finally to Wisdom?
- What is Knowledge Discovery?
- What are the goals of Knowledge Discovery?
- Explain the process of Knowledge Discovery.
- Explain Boxplots as a tool for Data Analysis.
- Why do we need a "Data Cleaning" step in Knowledge Mining?