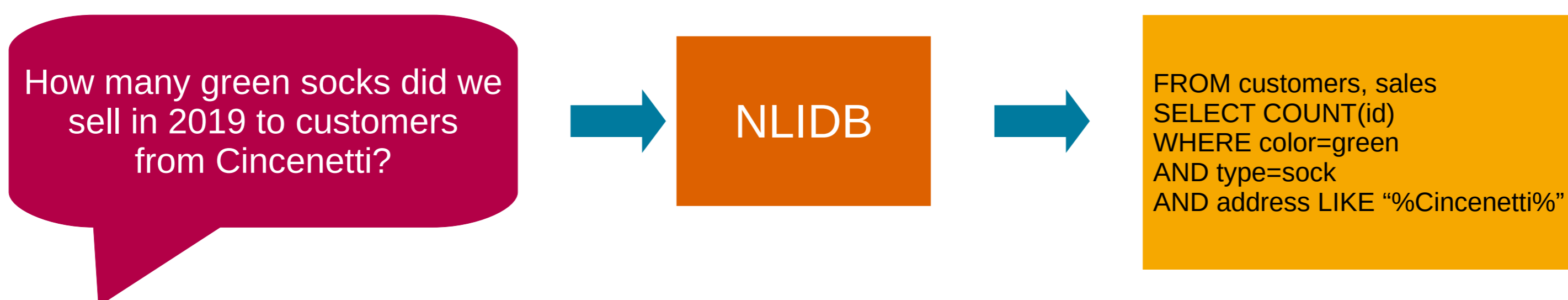# Low-Skill End User Data Analysis

Based on the Lecture by Carsten Binning

Data has increasingly become the fabric that connects our lives. Classical data analysis workflows feature a database expert assembling queries according to the domain expert's needs. As data becomes more omnipresent and fast-paced that approach starts to fall apart due to it's high turn-arount-time and cost.

Natural Language Interfaces to DBMS (NLIDB) take an unstructured query as a user might pose it to the database expert and try to derive an SQL (or similar) statement that acurately describes the query to the DBMS for execution.

How many green socks did we sell in 2019 to customers from Cincenetti?

→ NLIDB →

```
FROM customers, sales
SELECT COUNT(id)
WHERE color=green
AND type=sock
AND address LIKE "%Cincenetti%"
```

In addition, natural language interfaces also need to supply tools to explore the datasets. (E. g. find out which relations and which fields are available.) Good interfaces also guide the users and explain available information. More challenges arise when the output of the query also needs to be presented as natural language because complex and long tables and data structures are hard for humans to comprehend by hearing.

Voice assistens and services like WolframAlpha provide a basic version of NLIDBs. Notheless the area is actively researched. Commercial mainstream applications to explore custom datasets are not generally available.

Visual Interfaces to DBMS display the data in visual primitives that can be manipulated (for example by touch or click) into more complicated analysis and queries. Users can explore the dataset(s) and directly interact with them. Traditionally visualisations took a given query and made them accessible to consumers. Visual database interfaces enable users to create such queries **and** explore them.

Visual interfaces are an ongoing area of research. Mainstream commercial applications are not available.
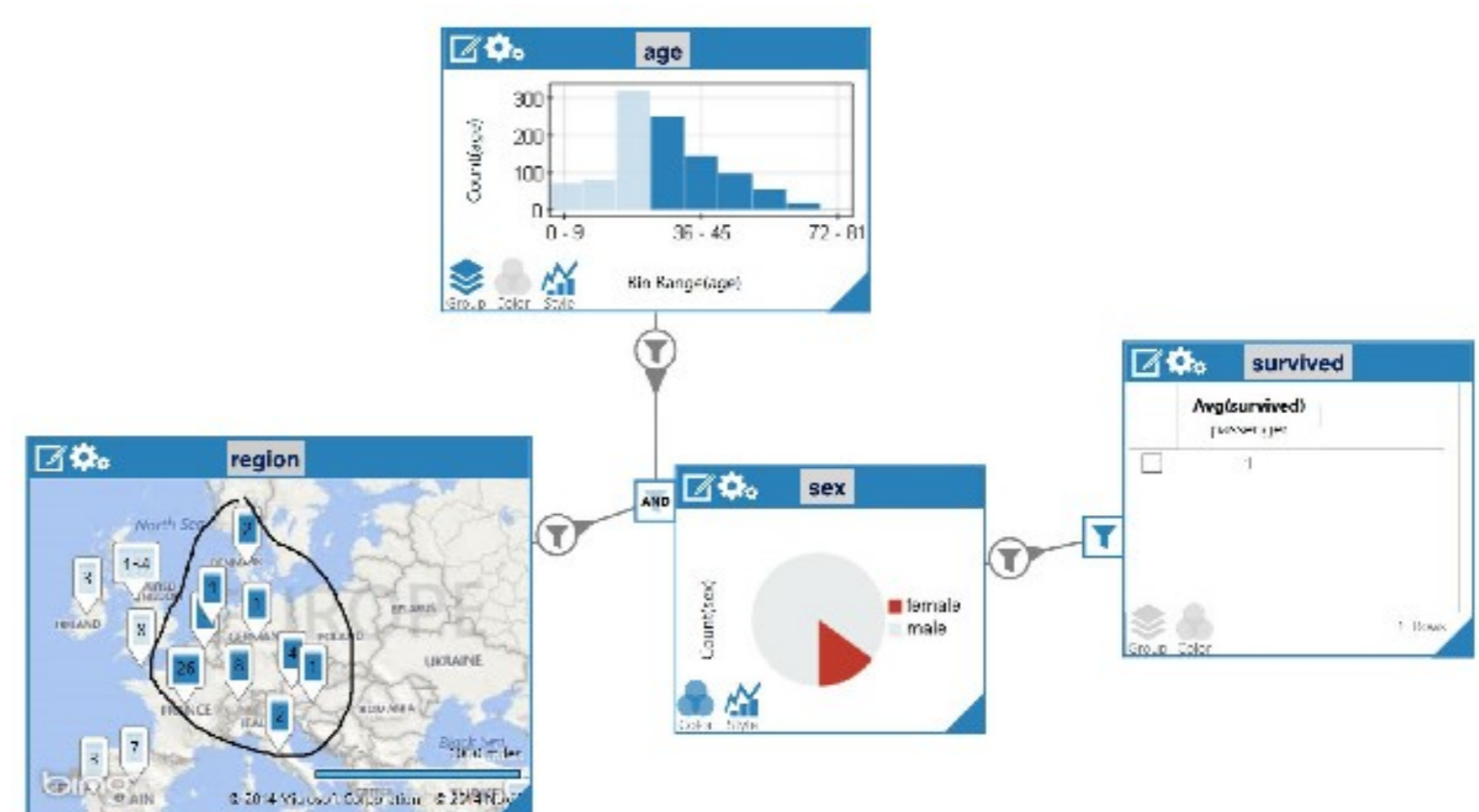
Example vizdom Query, Carsten Binning, TU Darmstadt

Challenges

SPEED – Visual Applications require low latency to appear interactive, new cache mechanisms have to be developed. Alternatively statistical approximations can be used to give good enough but fast results.

PARSING – Natural language queries are flexible and aim to adapt to the user. That is their weeknees. They can only ever support a subset of possible inputs, iritating the users or giving wrong results.

Advantages

NO DATA EXPERT NEEDED – Traditional data science workflows rely on a database expert interacting with the domain expert. Intuitive approaches allow the user to analyse data on her own.

KNOWN INTERFACE ABSTRACTIONS – Typical users are used to speech and visual interfaces. Query languages on the other hand have a steep learing curve.

Ben Bals, ITSE BA

Lecure Series: Practical Data Engineering
Prof. Felix Naumann
Prof. Tilmann Rabl
Hasso Plattner Institute, Potsdam, Germany

E-Mail: ben.bals@student.hpi.de

HPI Hasso Plattner Institut