

Übungsblatt 2 - Informationsvisualisierung und Visual Analytics

Nick Martin, Tim Lukas

Aufgabe 1: Daten- und Task-Analyse

What? (Datenanalyse)

Bei den gegebenen Rohdaten handelt es sich um eine zweidimensionale Tabelle. Die Items (Zeilen dieser Tabelle) sind sechs individuelle Personen identifiziert von A bis F, wobei jede Person eine eindeutige Entität darstellt. Diese Personen werden durch fünf Attribute (Spalten) beschrieben: dabei handelt es sich um die Programmiersprachen Java, Perl, JavaScript, C#, Python. In den Zellen der Tabelle werden die jeweiligen Werte angegeben, welche die Selbsteinschätzung der jeweiligen Person für eine spezifische Programmiersprache repräsentieren. Diese Werte liegen auf einer Skala von 1 bis 5, bei diesen Werten gilt: Je höher der Wert, desto besser die Selbsteinschätzung in der jeweiligen Programmiersprache.

Attribut Typen die Attribute der Tabelle entsprechen folgenden Klassifikationen:

- **Personen:** Personen sind kategorische Attribute. Sie haben keine implizite Ordnung und dienen in dieser Tabelle als eindeutiger Schlüssel, um auf die Werte eines Items zuzugreifen.
- **Programmiersprachen:** Auch die Programmiersprachen sind kategorische Attribute.
- **Werte der Selbsteinschätzung:** Die Selbsteinschätzungen sind ordinale Attribute. Es existiert eine klare Rangfolge (5 ist besser als 4), die Abstände zwischen den Werten sind jedoch weder quantifizierbar noch notwendigerweise gleichmäßig.

Why? (Task)

Ziel: Die gestellte Aufgabe ("Wer ist der bester und wer der schlechteste Programmierer"), bei der es sich um eine Domänenfrage handelt in allgemeine Aktionen (Actions) und deren Ziele (Targets) zerlegen.

Ziele (Targets)

- **Attributes -> Extremwerte:** Das primäre Ziel ist es, die Extremwerte in den Selbsteinschätzungen zu ermitteln. Es wird nach den Personen mit der höchsten Gesamteinschätzung (Maximum) und der Person mit der niedrigsten (Minimum) gesucht.

Aktionen (Actions): Um jedoch die Extremwerte verwenden zu können, müssen die verschiedenen Attribute (Programmiersprachen) zu einem repräsentativen Wert, der die Gesamtfähigkeiten wiederspiegelt zusammengefasst werden. Dieses Vorgehen kann folgendermaßen beschrieben werden:

- **Produce:** Um den besten/schlechtesten Programmierer über alle Attribute zu ermitteln, muss ein neuer Wert ergänzt (produced) werden, der einen repräsentativen Gesamtwert darstellt. Dieser Wert kann aus den vorhandenen Daten abgeleitet werden (Derive). Eine Möglichkeit hierfür wäre das Summieren der Werte.
- **Search:** Um Anschließend basierend auf diesem neu erstellten Wert den besten/schlechtesten Programmierer müssen nur die Extremwerte ermittelt werden (Suche mit bekanntem Ort und bekanntem Ziel).
- **Identify:** Nach dem Vergleich müssen die Personen identifiziert werden, die die Kriterien "beste" und "schlechteste" erfüllen (Extremwerte).

Es ist also notwendig einen neuen Wert zu ergänzen (Produce), dieser Wert kann aus den vorhanden Daten abgeleitet werden (Derive)

Aufgabe 3 - Erklärung zur Visualisierung

Die Visualisierung zeigt den Zusammenhang zwischen dem durchschnittlichen Tageseinkommen (x-Achse) und der durchschnittlichen Kinderzahl pro Frau (y-Achse) in verschiedenen Ländern. Die Größe der Kreise repräsentiert die jeweilige Bevölkerungszahl, während die Farbe die Zugehörigkeit zu einem Kontinent anzeigt.

Es ist ein deutlicher negativer Zusammenhang erkennbar: Mit steigendem Einkommen sinkt die durchschnittliche Kinderzahl pro Frau. Länder mit niedrigerem Einkommen weisen tendenziell höhere Geburtenraten auf. Dieser Zusammenhang wird besonders im Vergleich zwischen den Kontinenten Europa und Afrika deutlich. Während der Großteil der afrikanischen Länder eine hohe Geburtenrate bei geringem Durchschnittseinkommen aufweist, zeigt sich in den europäischen Ländern ein gegenteiliges Bild: Die meisten verfügen über höhere Einkommen und niedrigere Geburtenraten.

Die Visualisierung eignet sich folglich, um die Zusammenhänge zwischen der ökonomischen Situation eines Landes bzw. dessen Einwohnern und dessen demografischer Entwicklung aufzuzeigen. Zudem ermöglicht sie einen direkten Vergleich zwischen den einzelnen Ländern und Kontinenten.