

Aufgabe 2.1 | Multivariate Daten¹

In diesem Abschnitt werden Vor- und Nachteile sowie die Qualifikation und Disqualifikation verschiedener Daten für einige Visualisierungstechniken (i.e. Parallel Coordinates, Scatter Plots, Parallel Sets, Star Plot) beschrieben.

Parallel Coordinates

Die Visualisierungstechnik Parallel Coordinates (geometrische Visualisierungstechnik) verwendet pro Dimension je eine eigene vertikale Linie. Mehrere dieser sind dann parallel mit einem Abstand anzuordnen. Konkrete Datensätze sind dann quer zu den vertikalen Linien verlaufende Linien. Diese schneiden sich an der konkreten Wertaussprägung mit den vertikalen Linien. Ein Beispiel für Parallel Coordinates ist in Abbildung 1 zu sehen.

Parallel Coordinates sind besonders gut zum Erkennen von Korrelationen zwischen Variablen. Eine positive Korrelation äußert sich in horizontalen Linien, eine negative Korrelation in gekreuzten Linien zwischen zwei Dimensionen.

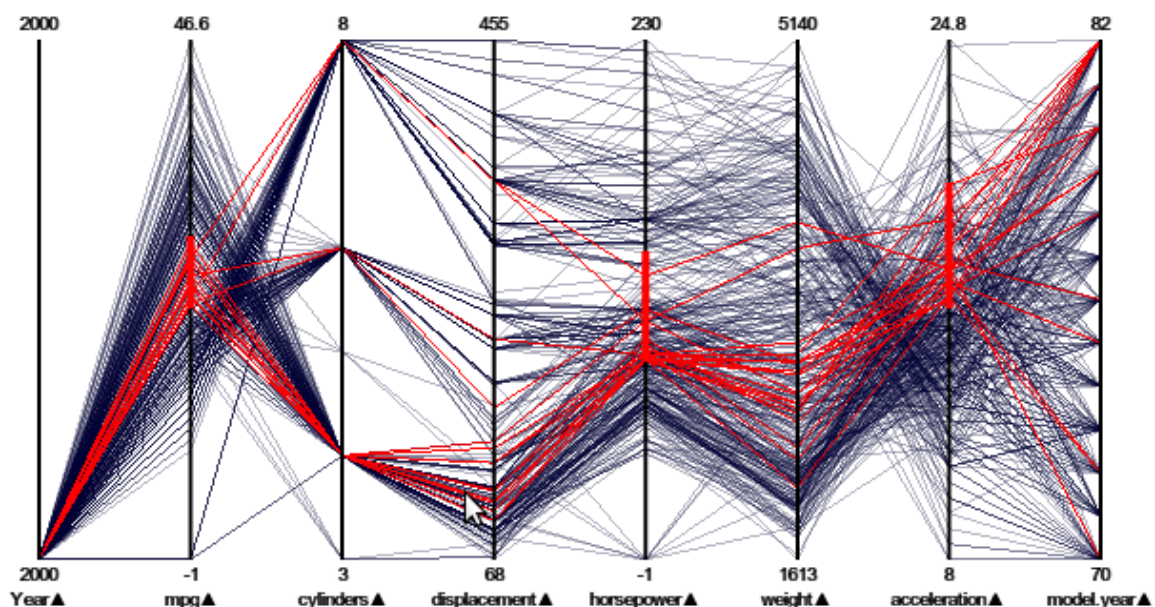


Abbildung 1: Visualisierung mit Hilfe von Parallel Coordinates

- **Vorteile / Daten die diese Methode qualifizieren:** Im Gegensatz zu Scatter Plots auch bei mehr als drei Dimensionen verwendbar. Quantitative (numerische), ordinale und nominale Daten können verwendet werden.
- **Nachteile / Daten die diese Methode disqualifizieren:** Bei zu vielen Datenpunkten sollte eingegriffen werden. Als Möglichkeiten sind hier folgende Möglichkeiten anzuden-

¹Die Qualität der Bilder die man zum Thema Visualisierung im Internet findet lässt oft zu wünschen übrig. Weiterhin werden oft Bilder die als „png“ gespeichert werden sollten, als „jpg/jpeg“ angeboten.

ken: Brushing, das Eingrenzen von Wertebereichen, das Vertauschen von Achsen oder verschiedene Clustertechniken wie das hierarchische Clustering.

Scatter Plot

Ein Scatter Plot (geometrische Visualisierungstechnik) beschreibt Datenpunkte, eingetragen in ein 1-,2- oder 3-dimensionales Koordinatensystem — so wie man es aus der Mathematik kennt. Ein Beispiel für ein Scatter Plot ist in Abbildung 2 zu sehen.

Scatter Plots sind besonders gut zum Erkennen von Objektclustern. Cluster erkennt man optisch trivial an einer erhöhten Anzahl von Objekten in einem bestimmten Bereich im Koordinatensystem.

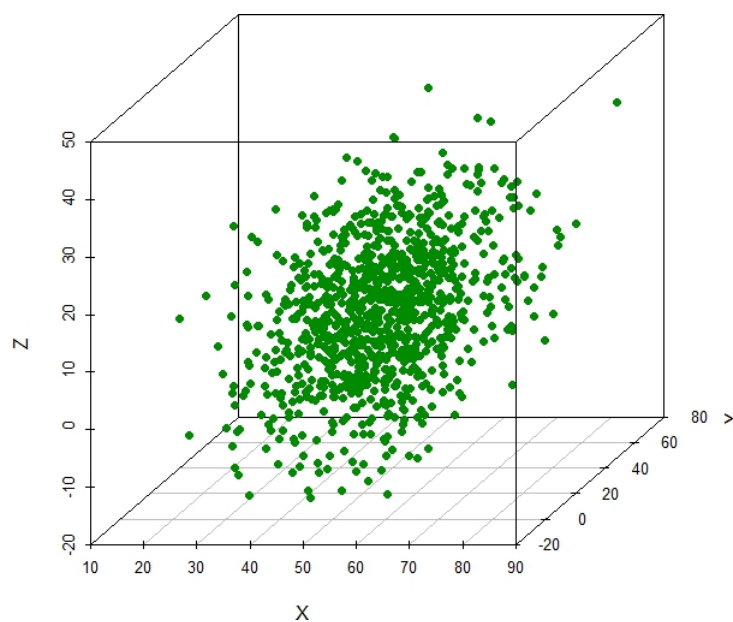


Abbildung 2: Scatter Plot mit drei Dimensionen

- **Vorteile / Daten die diese Methode qualifizieren:** Scatter Plots sind gut für eine schnelle Übersicht, zur Strukturfindung und um Zusammenhänge zu erkennen. Daten mit einer, zwei oder maximal drei Dimensionen. Große Anzahl an Datenpunkten möglich. Bei zu vielen Datenpunkten sollten mehrere Datenpunkte zu einem zusammengefasst werden (Diese Eigenschaft haben alle geometrischen Visualisierungstechniken gemeinsam). Quantitative (numerische), ordinale und nominale Daten können verwendet werden.
- **Nachteile / Daten die diese Methode disqualifizieren:** Eine hohe Dimensionalität des Datensatzes. Da Scatter Plots in einem kartesischen Koordinatensystem dargestellt werden, können nur 1 und 2 (bzw. begrenzt 3) Dimensionen dargestellt werden. Die dritte Dimension ist vom Betrachtungsmedium abhängig besser (Computer, interaktiv) oder schlechter (Papier, Computer, nicht-interaktiv) interpretierbar.

Parallel Sets

Parallel Sets ist eine geometrische Visualisierungstechnik. Eine horizontale Linie, unterteilt in verschiedene Abschnitte (je nach Ausprägungen der ersten Dimension) wird erstellt. Die verschiedenen Abschnitte gehen nun in unterschiedlicher Breite zur nächsten Dimension über. Am besten erkennt man das ganze an einem Beispiel (Abbildung 3). Man erkennt z.B., dass etwa 60% des harten Wassers („hard“) bei einer geringen Temperatur („Low“) getrunken wird. Die restlichen 40% des harten Wassers werden bei hoher Temperatur getrunken („High“).

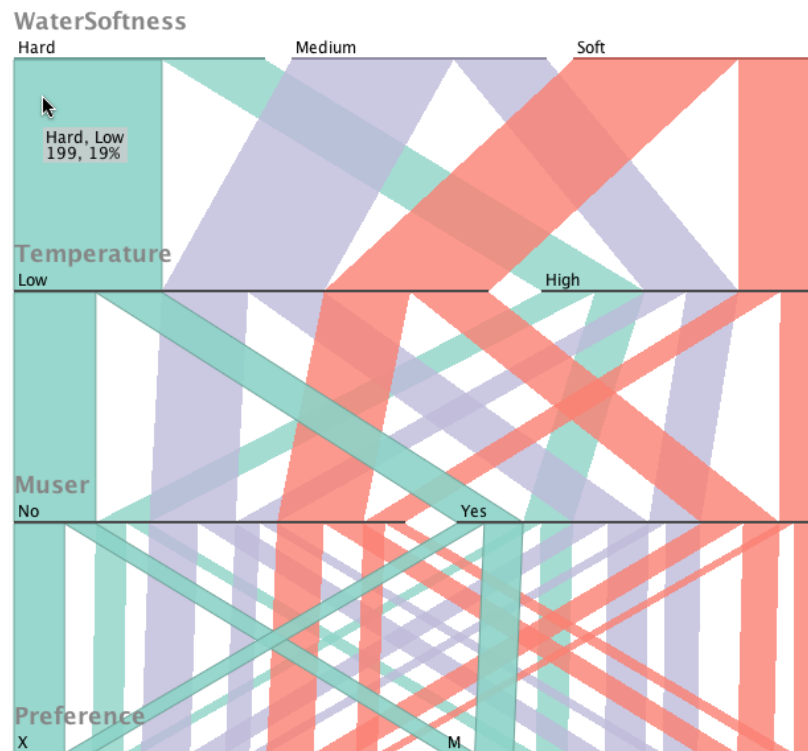


Abbildung 3: Ein Parallel Set als Beispiel

- **Vorteile / Daten die diese Methode qualifizieren:** Die Anzahl der möglichen Dimensionen ist theoretisch nicht begrenzt, allerdings verliert die Darstellung an Übersicht wenn zu viele Dimensionen verwendet werden. Auch hier ist es vom Medium abhängig wie gut eine hohe Dimensionsanzahl vom Menschen verarbeitet werden kann. Hovereffekte, die einen bestimmten Teilbaum eines Parallel Sets hervorheben könnten zum Beispiel vom Vorteil sein. Für Parallel Sets können quantitative (numerische), ordinale und nominale Daten verwendet werden.
- **Nachteile / Daten die diese Methode disqualifizieren:** Wie in den Vorteilen schon angesprochen wirkt sich eine hohe Dimensionsanzahl negativ auf die Übersichtlichkeit aus.

Star Plot

Der Star Plot (manchmal auch Star Glyph) (icon-basierte Visualisierungstechnik) besteht aus einem Mittelpunkt, einer Linie für jede Datendimension (angeordnet im Kreis) sowie schließlich jeweils eine Linie zwischen den Datendimensionenlinien. Somit entsteht eine einem Spinnennetz ähnliche Struktur, ein Polygon. Ein Beispiel (hier mit drei durch Farben getrennte Datensätze) kann in Abbildung 4 betrachtet werden.

Der Star Plot ist besonders gut zum Erkennen von Ausreißern geeignet. Das rührt daher, zum einen zwischen Star Plots eine gute Vergleichbarkeit herrscht. Zum anderen stechen Ausreißer durch eine sehr steile Spitze (positiver Ausreißer) oder durch eine fast nicht sichtbare Dimensionsausprägung (negativer Ausreißer) hervor.

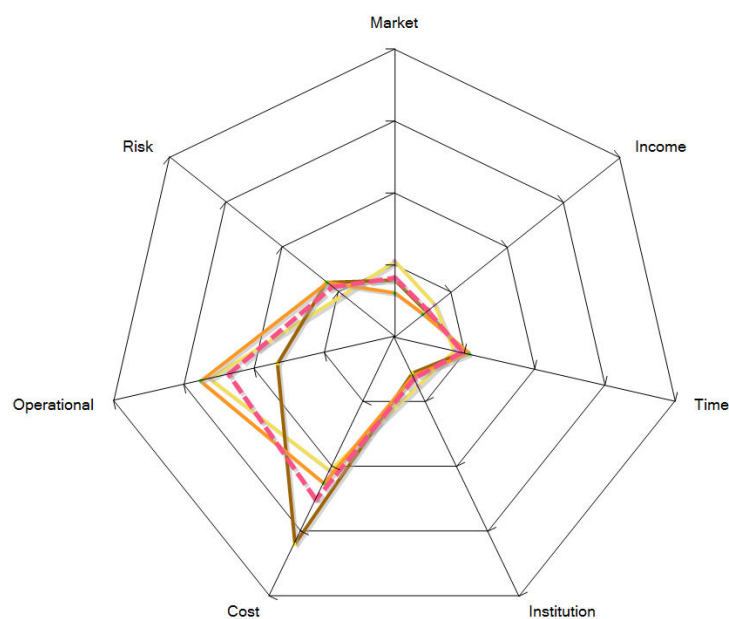


Abbildung 4: Ein Parallel Set als Beispiel

- **Vorteile / Daten die diese Methode qualifizieren:** Durch die geometrische Anordnung der Ausprägungen ist es schnell möglich, Vergleiche zwischen verschiedenen Datensätzen auszumachen.
- **Nachteile / Daten die diese Methode disqualifizieren:** Wie auch bei den geometrischen Visualisierungstechniken ist auch hier die Skalierbarkeit ein Problem. Existieren viele Dimensionen so ist das Argument der Vergleichbarkeit zwischen verschiedenen Datensätzen nicht mehr gegeben. Weiterhin sind Icon-basierte Visualisierungstechniken, also auch der Star Plot, nur nützlich wenn qualitative (und nicht quantitative) Aspekte von Datensätzen dargestellt werden sollen. Das kommt daher, dass pro Dimension effektiv nur eine sehr geringe Anzahl an Ausprägungen übersichtlich dargestellt werden kann.

Aufgabe 2.2 | RadViz

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: Dies ist ein Blindtext oder Huardest gefburn? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie Lorem ipsum dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

Aufgabe 2.3 | Visualisierung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: Dies ist ein Blindtext oder Huardest gefburn? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muß keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie Lorem ipsum dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.