

Wolfram Fischer

Neue Grafiken zur Datenvisualisierung

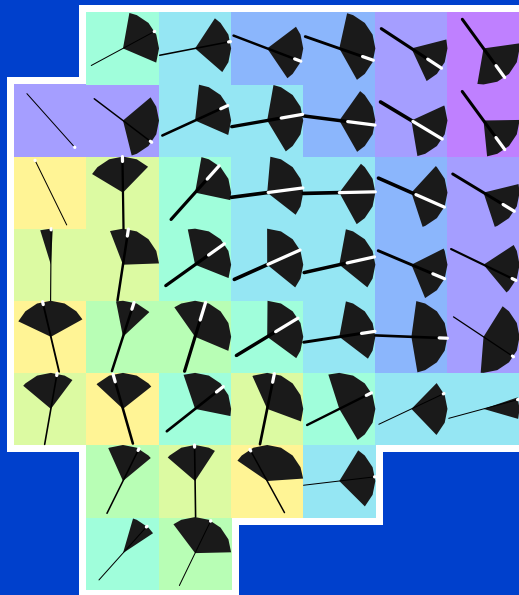
Band 1: Speichengrafiken, Streuungsfächerkarten,
Differenz-, Sequenz- und Wechseldiagramme

Wolfram Fischer

Neue Grafiken zur Datenvisualisierung

Band 1

Speichengrafiken, Streuungsfächerkarten,
Differenz-, Sequenz- und Wechseldiagramme



Zentrum für Informatik und wirtschaftliche Medizin
Wolfertswil



Angaben zum Autor

Wolfram Fischer
ZIM – Zentrum für Informatik und wirtschaftliche Medizin
Steigstrasse 12
CH-9116 Wolfertswil (SG)
wolfram@fischer-zim.ch

Betriebswirtschaftler (lic. oec. HSG) und Medizin-Informatiker;
Gründer und Leiter des ZIM.

Tätigkeitsschwerpunkte:

- Analyse und Visualisierung medizinökonomischer Daten.
- Patientenklassifikationssysteme aus ärztlicher und aus pflegerischer Sicht.
- Bereitstellung fachspezifischer Informationen im Internet unter:
www.fischer-zim.ch

Liste der «Kunstgrafiken» in diesem Buch

- Mond-Feuer-Blume (2010) ↑ S. 17
- Linien-Aussaat (2009) ↑ S. 29
- Dreiecke im Wind (2005) ↑ S. 45
- Rot verloren im Blauwald (2010) ↑ S. 71
- Wechselflaggen-Gehänge (2010) ↑ S. 85
- Speichenspirale 1 (2009) ↑ S. 97

Informationsseite zum Buch

<http://www.fischer-zim.ch/studien/Neue-Grafiken-I-1003-Info.htm>

© 2010 (Internetversion) Wolfram Fischer, Wolfertswil
Zentrum für Informatik und wirtschaftliche Medizin
<http://www.fischer-zim.ch/verlag/>
Umschlag, Layout und Satz: Wolfram Fischer
(Verwendete freie Software: \LaTeX , m4, R, perl, etc. unter Linux)
Grafik auf Umschlag: Auf der Grundlage von Tafel 8, S. 39
Korrektorat: Edith Meier-Keim

ISBN 978-3-905764-06-2

Vorwort

Statistische Grafiken

Statistische Grafiken helfen, Zahlenmaterial besser zu verstehen. Sie lassen nicht nur Besonderheiten, sondern auch Unauffälligkeiten erkennen. Sie sollen den Leser unterstützen, Fragen zu stellen und zu diskutieren.

Plakativ oder zisiert?

Es gibt zwei grundsätzlich verschiedene Ansätze, Grafiken zu gestalten: Mit einer plakativen Grafik kann man wirkungsvoll eine bestimmte Aussage belegen. Man destilliert dazu aus einer Vielzahl von vorliegenden Informationen jene Daten heraus, deren Darstellung den *Betrachter der Grafik* von der angestrebten Aussage überzeugt.

Mit einer differenzierten, dichten Grafik kann man einen Sachverhalt umfassender darstellen. Auch dazu wird man Informationen auswählen müssen. Man zeigt aber soviel Information, dass der *Leser der Grafik* selbst einen Schritt in die Tiefe machen kann und dass er möglicherweise Fragen an die Auswertungen stellt, die vom Konstrukteur der Grafik so nicht vorausgedacht worden waren.

Kreative Konstruktionen

Aussagekräftige und diskussionsfördernde statistische Grafiken zu gestalten ist eine herausfordernde und zugleich auch sehr spannende, ja beflügelnde Aufgabe. Wenn man die Freiheit hat, Grafiken jenseits der Standardvorschläge aus einem Officepaket zu gestalten, beginnt die Arbeit – nach der vorgängigen Präzisierung des Darstellungsziels – zuerst einmal mit dem Sammeln von Gestaltungsideen: Welche Möglichkeiten gäbe es, mit den verschiedenen Aspekten des Datenmaterials umzugehen? Genügen die bekannten Darstellungstechniken oder könnte eine neue Form entwickelt werden? Bei der anschliessenden Umsetzung und Erprobung der Ideen werden sich neue Wege auf tun; andere Wege werden sich als ungeeignet entpuppen. Schliesslich entscheidet man sich für eine bestimmte Ausgestaltung, die man als gelungen empfindet. Oder man verwirft das Ganze, nimmt eine andere Idee auf und beginnt von vorne.

Entdeckungsreise

Auf den folgenden Seiten stelle ich einige der von mir auf solchen Entdeckungsreisen entwickelten neuartigen Grafikelemente und Grafiken vor.

Spielfreude

Dass diese Arbeit auch zum Spiel an- oder verleitet, zeigen die den Kapiteln vorangestellten künstlerischen Verfremdungen der Grafiken. Sie beruhen auf realem Zahlenmaterial, wurden aber durch gezielte – manchmal zunächst unabsichtliche – Fehleinstellungen einzelner Parameter umgeformt.

Inhaltsverzeichnis

Tabellen und Abbildungen	9
A Einleitung	11
A.1 Aufbau des Buches	11
A.2 Lesehinweise	12
A.3 Vorüberlegungen	13
B Speichengrafiken	15
B.1 Motivation	19
B.2 Beschreibung	20
B.3 Beispiel 1: Überblick über Fallzahlen in 445 Patientenkate- gorien	21
B.4 Beispiel 2: Preisunterschiede in 546 Patientenkategorien . .	24
C Streuungsfächerkarten	27
C.1 Motivation	31
C.2 Beschreibung: Fächer	32
C.3 Beispiel 1: Fähigkeitseinschränkungen und Pflege	34
C.4 Beispiel 2: Fähigkeitseinschränkungen und Pflege im Zeit- verlauf	36
C.5 Beispiel 3: Mathematikprüfungen und sozioökonomischer Status	38
C.6 Beispiel 4: Krankenkassenprämien	40
D Differenzdiagramme	43
D.1 Motivation	47
D.2 Beschreibung: Dreiecke	47
D.3 Beispiel 1: Vergleich von Fallanteilen in Patientenkategorien	48
D.4 Beispiel 2: Gerstenernten 1931 und 1932	50
D.5 Beschreibung: Punktbalken	52
D.6 Beispiel 3: Gerstenernten 1931 und 1932	52
D.7 Vergleich von vier Darstellungsweisen	54
D.8 Beschreibung: Dreieck mit Balken	56
D.9 Beispiel 4: Gerstenernten 1931 und 1932	56
D.10 Beschreibung: Kombination von zwei Dreiecken	58
D.11 Beispiel 5: Gerstenernten 1931 und 1932	58

E Sequenzdiagramme	69
E.1 Motivation	73
E.2 Beschreibung: Sequenzierte Dreiecke	75
E.3 Beispiel 1: Verlauf von Rehabilitationsbehandlungen	76
E.4 Beschreibung: Sequenzbänder	78
E.5 Beispiel 2: Verlauf von Rehabilitationsbehandlungen	78
E.6 Beschreibung: Kammlinien	80
E.7 Beispiel 3: Verlauf von Rehabilitationsbehandlungen	80
F Wechseldiagramme	83
F.1 Motivation	87
F.2 Beschreibung: Wechselflagge	88
F.3 Beispiel 1: Antwortpaare zu Ja/Nein-Fragen nach Ländern	88
F.4 Beispiel 2: Antwortpaare zu Ja/Nein-Fragen nach Themen	90
F.5 Beispiel 3: Antwortkombinationen zu Ja/Nein-Fragen pro Person	92
G Anhang	95
G.1 Statistische Anmerkungen	99
G.1.1 Median und Quartile in Boxplot und Streuungsfächerkarte	99
G.1.2 Varianzreduktion (r^2)	100
G.1.3 Reduktion der absoluten Abweichungen vom Median (r_1)	102
G.2 Verzeichnisse	104
G.2.1 Abkürzungen	104
G.2.2 Literaturverzeichnis	105
— Stichwortverzeichnis	107

Tabellen und Abbildungen

1	Platzbedarf einer Speichengrafik im Vergleich zum Balkendiagramm	21
2	SGP 2005: Anzahl Fälle in allen belegten APDRGs	23
3	HRG 3.5: Unterschiede der HRG-Pauschalen 2008 für elektive und nicht-elektive Behandlungen	25
4	Erster Versuch zu einem grafischen Symbol für ein Strömungsbild	31
5	Legende zum Streuungsfächer (360°)	33
6	Median und Quartile der LEP-Stunden pro Tag, nach motorischem und kognitivem FIM (TAR97)	35
7	Median und Quartile der LEP-Stunden pro Tag, nach motorischem und kognitivem FIM (Test-Daten)	37
8	7185 Mathematikprüfungen: Resultate nach Geschlecht und Zugehörigkeit zu einer Minorität	39
9	CH: Krankenkassenprämienangebote	41
10	Legende zum Symbol für Differenzdiagramme	47
11	Krankenhaus H217: Fallanteile nach APDRG-Subkategorien und Kostengewichtsklassen im Vergleich zu allen Krankenhäusern des Typs «Zentrumsversorgung»	49
12	Gerstenernten in Minnesota 1931 und 1932 (Differenzdiagramm)	51
13	Legende zum Symbol für Punktbalkendiagramme	52
14	Gerstenernten in Minnesota 1931 und 1932 (Punktbalkendiagramm)	53
15	Gerstenernten in Minnesota 1931 und 1932 (vier Darstellungsweisen)	55
16	Legende zum Symbol für Doppeldifferenzdiagramme	56
17	Gerstenernten in Minnesota 1931 und 1932 (Doppeldifferenzdiagramm)	57
18	Legende zum Symbol für Differenzdiagramme mit zwei Dreiecken	58
19	Gerstenernten in Minnesota 1931 und 1932 (zwei Differenzdiagramme)	59
20	Gerstenernten in Minnesota 1931 und 1932, korrigiert (zwei Differenzdiagramme)	61
21	Gerstenernten in Minnesota 1931 und 1932, korrigiert (zwei Differenzdiagramme), nach Sorten	63
22	Gerstenernten in Minnesota 1931 und 1932, korrigiert (zwei Differenzdiagramme), nach Anbauorten	65
23	Gerstenernten in Minnesota 1931 und 1932, korrigiert (Strömungsbilder)	67
24	Legende zu sequenzierten Dreiecken mit fünf Werten	73
25	Legende zu sequenzierten Dreiecken mit sieben Werten	74
26	Legende zu sequenzierten Dreiecken mit acht Werten	75
27	Entwicklung der FIMø-Punkte nach Hospitalisationswochen	77
28	Legende zu einem Sequenzband mit acht Werten	78
29	Sequenzbänder zur Entwicklung der FIMø-Punkte nach Hospitalisationswochen	79
30	Legende zu einer Kammlinie mit acht Werten	80
31	Kammlinien zur Entwicklung der FIMø-Punkte nach Hospitalisationswochen	81
32	Legende zu den Wechselflaggen	87
33	Geordnete Antwortpaare zu sieben Ja/Nein-Fragen, nach Ländern	89
34	Geordnete Antwortpaare zu sieben Ja/Nein-Fragen, nach Themen	91

35	Antwortpaare pro Person zu sieben Ja/Nein-Fragen, nach Ländern und Personen	93
36	Beispiel zur Berechnung der Varianzreduktion	101
37	Beispiel zur Berechnung der Reduktion der absoluten Abweichungen vom Median	103
38	Im Text verwendete Abkürzungen	104

A Einleitung

A.1 Aufbau des Buches

Grafiken gestalten	Während meiner langjährigen Wühlarbeit in Datenbergen aus dem Gesundheitswesen begann ich, eigene Grafiken zu gestalten. Dabei kamen mir meine Programmierkenntnisse zugute, und die Programmiersprache «R» ¹ eröffnete mir ganz neue Möglichkeiten. Sie spornte mich an, grafische Darstellungen von Grund auf neu zu denken.
Buchreihe	In der mit diesem Band eröffneten Buchreihe werden von mir entwickelte neuartige Typen von Grafiken zur Visualisierung statistischer Daten vorgestellt. Sie dienen zur differenzierten Darstellung umfangreicher und komplexer Daten. Und sie sollen mit einer ansprechenden Gestaltung auch das Auge erfreuen.
Kapitelinhalte	Kapitelweise werden folgende Typen von Grafiken vorgestellt: <ul style="list-style-type: none"> • <i>Speichengrafik</i>²: Balkendiagramm in Kreisform. • <i>Streuungsflächenkarte</i>: Fächersymbole mit fixer Grösse, mit welchen die gleichen Verteilungskennzahlen dargestellt werden können wie mit Boxplots, mit Vergleichsmöglichkeiten in zwei Dimensionen. • <i>Differenzdiagramme</i>³: Dreiecke – z. T. mit Balken kombiniert – zur Anzeige von Paaren oder Tripeln von absoluten Werten und deren Differenzen. • <i>Sequenzdiagramme</i>⁴: Dreiecke und Bänder zur Anzeige von sequenzierten absoluten Werten und deren Differenzen. • <i>Wechseldiagramm</i>: Diagonal unterteilte Quadrate zur Darstellung von zwei Zuständen und zwei Zustandswechseln.
«Kunstgrafiken»	Jedem Kapitel wurde eine «Kunstgrafik» vorangestellt. Sie wurden mit den gleichen Algorithmen und Daten wie die statistischen Grafiken konstruiert, aber mit anderen Werten parametrisiert, um schöne oder überraschende, abstrakte Grafiken zu erhalten.
Statistische Fachausdrücke	Im Anhang folgen einige Anmerkungen zu statistischen Fachausdrücken, welche im Rahmen der Präsentation der Streuungsflächenkarten

↑ S. 15

↑ S. 27

↑ S. 43

↑ S. 69

↑ S. 83

↑ S. 99

¹ [http:// www.r-project.org /](http://www.r-project.org/); Chambers [R, 2008]; Dalgaard [R, 2002]. – «R» ist eine OpenSource-Implementation der Programmiersprache «S»; vgl. dazu: Chambers [S, 1998].

² Erstmalige Beschreibung in: Fischer [Grafiken zur PCS-Beurteilung, 2008]: 35.

³ Erstmalige Beschreibung in: Fischer [Krankenhaus-Betriebsvergleiche, 2005]: 135 ff.

⁴ Erstmalige Beschreibung in: Fischer [Krankenhaus-Betriebsvergleiche, 2005]: 108 f.

verwendet wurden. Sie sind für Leser gedacht, denen diese Dinge noch nicht besonders geläufig sind.

Im Text anzutreffende Abkürzungen sind ebenfalls im Anhang zusammengestellt und mit allfälligen Internetverweisen ergänzt. Anschliessend folgen Literatur- und Stichwortverzeichnis.

A.2 Lesehinweise

In diesem Buch gibt es zwei Arten von internen Querverweisen:

- Verweise auf Tafeln, die unmittelbar zum Text gehören.
- Verweise auf Tafeln oder Abschnitte, die in anderen Teilen des Buches zu finden sind.

In der Pdf-Datei gibt es in den Marginalien und im Inhaltsverzeichnis blau eingefärbte Tabellen- und Seitennummern. Durch einen Mausklick gelangt man direkt zur referenzierten Seite.

Grau gedruckte Texte enthalten Verweise auf Texte im Internet, die für diese Arbeit gebraucht wurden. Durch einen Mausklick öffnet sich ein Internet-Browser-Fenster, und man kann sich den referenzierten Text ansehen. Da die Inhalte im Internet z. T. sehr schnelllebig sind, kann es sein, dass manche Verweise bereits zum Zeitpunkt des Lesens nicht mehr gültig sind. In solchen Fällen kann es sich lohnen, im Internet nach Titel und Autor zu suchen, um einen allfälligen neuen «Standort» ausfindig zu machen.

Bei mehrzeiligen Internetverweisen kann es sein, dass man die aktive Zone für den Mausklick kaum findet. (Sie befindet sich dann zwischen den Zeilen.) Dies schafft die aktuelle Version der verwendeten Textaufbereitungssoftware leider noch nicht besser.

Es ist eine der Spezialitäten der hier vorgestellten Konstruktionsweisen, dass Grafiken auch dann noch auf einer Seite Platz haben, wenn sie grosse Datenmengen abbilden. Das Ziel ist es, gleichzeitig sowohl einen Überblick zu erhalten als auch auffällige Details sichtbar zu machen. Ein Nachteil dieser Technik ist, dass die Texte in den resultierenden Grafiken zum Teil sehr klein geworden sind und dass es notwendig werden kann, die Lupe zu Hilfe zu nehmen.

Verzeichnisse
↑ S. 104

Querverweise:
► direkter Verweis
↑ Querverweis

Blaue Verweise auf
Tabellen und Seiten

Graue Verweise
ins Internet

Dichte Grafiken mit
z. T. kleinen Texten

Überblick
und Vergleiche

Relative und
absolute Werte

Lokalisierbarkeit

Relevanz

Erster und
zweiter Blick

Ansprechendes
Aussehen

Lernerfahrungen

Kleine Formen

Weniger
aggregierte Daten

A.3 Vorüberlegungen

Als generelle Hauptaufgaben einer Grafik sind mir wichtig:

1. Eine Grafik soll eine Einschätzung der Daten ermöglichen. Dazu ist es nötig, dass sie dem Leser einen Überblick verschafft und gleichzeitig verschiedene Vergleichsmöglichkeiten bereitstellt.
2. Vergleiche werden durch die Abbildung relativer Werte oder durch die Gegenüberstellung absoluter Werte ermöglicht.
3. Auch wenn relative Werte direkt abgebildet werden, sollen die absoluten Werte nach Möglichkeit erkennbar sein, damit die Bedeutsamkeit ihres Ausmasses abgeschätzt werden kann.
4. Die Daten müssen «lokalisiert» werden können: Sie müssen einer Kategorie (oder mehreren) und/oder einem Mass (z. B. Zeit, Koordinaten, identifizierender Wert) zugeordnet werden können.
5. Die Relevanz der präsentierten Kennzahlen soll soweit möglich abgeschätzt werden können. (Hilfreich dazu können z. B. Mengenangaben sein, welche als Indikatoren der Relevanz betrachtet werden können.)
6. Eine Grafik soll aus zwei Distanzen betrachtet werden können: Aus grösserer Entfernung soll man ein Gesamtbild erhalten und erkennen können, ob dieses eher ruhig ist oder ob es auffällige Bereiche gibt, die Anstoss zu Fragen geben könnten. Für eine nähere Betrachtung soll die Grafik mit genügend Detailinformationen ausgestattet sein, um ein facettenreicheres Bild zu ermöglichen und Hinweise zu Antworten auf einige der entstandenen Fragen liefern zu können.
7. Eine Grafik soll ansprechend aussehen.

Bei der Arbeit an Grafik-Konstruktionen und bei der Auseinandersetzung mit deren Konstruktionsprinzipien konnte ich viel lernen. Fasziniert war ich insbesondere von der Entdeckung, dass das menschliche Auge nicht nur auf plakativ präsentierte Unterschiede reagiert, sondern auch auf erstaunlich geringe Veränderungen von Farben, Formen und Positionen.

Aufgrund dieser Erkenntnis begann ich, mit kleineren Formen und mit Untergliederungen in «Grafikfelder» zu arbeiten. Es zeigte sich, dass die Vergrösserung von Grafiken eher selten eine wirklich differenziertere Interpretation ermöglichte. Eine Vergrösserung ist dann hilfreich, wenn nicht mehr alle Textelemente entziffert werden können. Da eine Vergrösserung gleichzeitig aber auch zu einem Verlust an Übersichtlichkeit führen kann, ist nach einem Optimum zwischen Kompaktheit und Lesbarkeit zu suchen.

Kleinere Formen erlauben es auch, Daten darzustellen, die weniger stark aggregiert sind. Dies gibt bessere Einsichten in Strukturen wie z. B. Verteilungsstrukturen.

Ich setzte nicht nur verschiedene Farben, sondern auch unterschiedliche Farbintensitäten und Farbverläufe ein. Damit war eine vielfältigere Farbgebung möglich. Darüberhinaus ergaben sich so mehr Codierungsmöglichkeiten. Und: Nicht-grelle Farben sind oft angenehmer zum Betrachten.

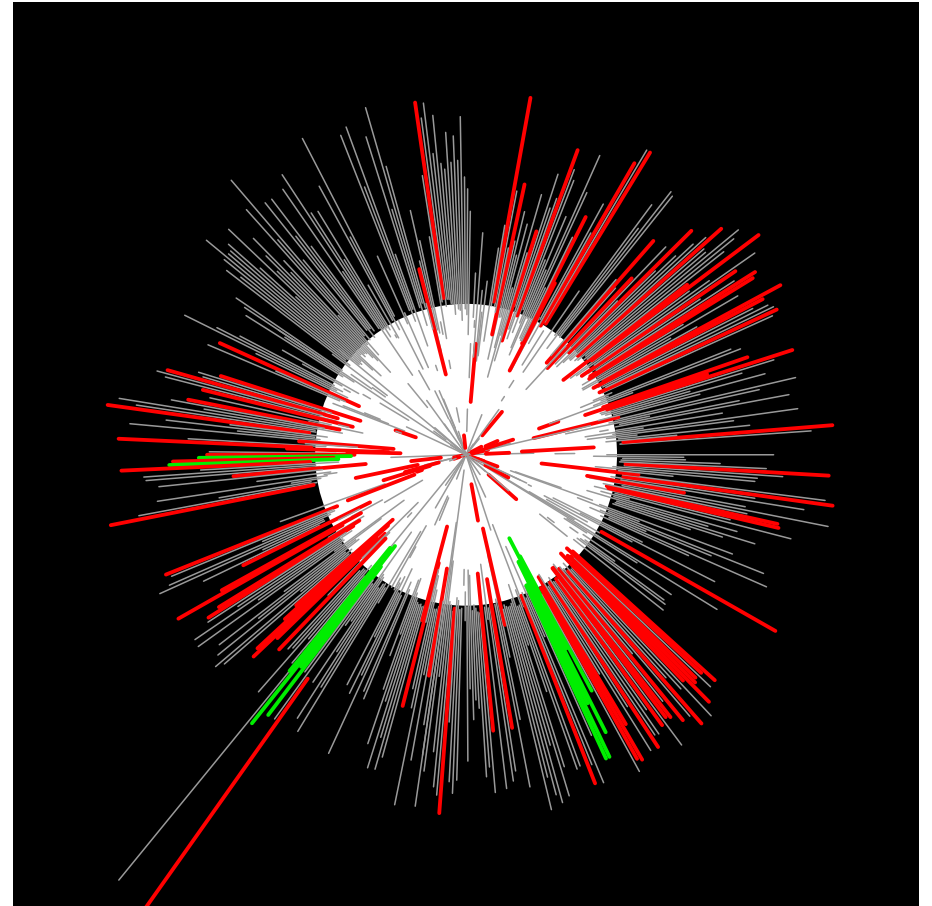
Bei dieser Konstruktionsarbeit erkannte ich im Weiteren, dass grafische Symbole mit fixer Ausdehnung für kompakte Darstellungen sehr dienlich sind. Ausserdem werden damit Vergleiche in zwei Dimensionen (senkrecht und waagrecht) möglich.

Nach diesen Vorüberlegungen möchte ich den Leser nun nicht mehr länger mit theoretischen Überlegungen aufhalten, sondern die Vorstellung der Grafiken mit den Speichengrafiken beginnen . . .

Farbintensitäten
und Farbverläufe

Grafiksymbole mit
fixer Ausdehnung

B Speichengrafiken



Mond-Feuer-Blume (2010)

Aus der Speichengrafik mit Differenzen zu gut 500 HRG-Pauschalen, ↑ Tafel 3 (S. 25)
durch beidseitige Verkürzung der radialen Skala.

B.1 Motivation

Lange Liste von Häufigkeiten	Was macht man am besten mit einer endlos langen Liste von Häufigkeiten? Das erste ist natürlich, dass man versucht, die Liste zu strukturieren. Man sucht nach Gruppen, nach einer hierarchischen Struktur. Solche Bündelungen helfen bei der Orientierung.
Balkendiagramm?	Trotzdem bleibt die Ausgangsliste lang, sofern man nicht auf eine detaillierte Darstellungsform verzichten kann. Eine konventionelle Darstellung mit Balken, welche die Häufigkeiten abbilden, kann an mangelndem Platz scheitern oder an den grossen Häufigkeitsunterschieden, welche schwach belegte Kategorien manchmal beinahe unsichtbar werden lassen. ¹
Durchmesser $\times \pi$ = Kreisumfang	Als ich an dieser Fragestellung herumstudierte, ging mir durch den Kopf, dass ein abgewickelter Kreis eine lange Linie erzeugt. Sie ist gut dreimal so lang wie der Durchmesser des Kreises. Was würde passieren, dachte ich, wenn ich den umgekehrten Vorgang benutzte? Wie wäre es, wenn ich die Grundlinie, auf der die Balken stehen, zu einem Kreis formte?
Mehrere hundert Kategorien in einem Kreis	Als ich dem Computer beigebracht hatte, die Balken in einen Kreis zu stellen und diese dann zum ersten Mal betrachtete, staunte ich, dass ich trotz der Darstellung der Häufigkeiten zu mehreren hundert Kategorien nicht den Eindruck hatte, von der Datenmenge überfahren zu werden, sondern dass ich – vielleicht auch angezogen von der schönen Form – interessiert schauen ging, wo viel und wo wenig Messungen vorlagen. Ich merkte, dass eine Gruppierung nach wie vor nötig war. Aber die Gruppen büschelten sich kompakt, und es resultierte ein überraschend übersichtlicher Gesamteindruck.
«Speichengrafik»	Die neue Grafik nannte ich «Speichengrafik», weil sie mich an ein Rad eines Fahrrades erinnerte.

¹ Von Hoffmann stammt der interessante Vorschlag, unsichtbare, aber trotzdem belegte Positionen bei Histogrammen oder anderen Balkendiagrammen mit einem kleinen (z.B. roten) Markierungsstrich zu versehen. – Vgl. Unwin et al. [Graphics, 2006]: 58+87+106.

B.2 Beschreibung

«Speichengrafiken» sind Balkendiagramme in Kreisform. Damit können auf einem beschränkten Platz ungefähr dreimal so viele Einträge gezeigt werden wie mit einem gewöhnlichen Balkendiagramm. Dank der Gewohnheit, tagtäglich analoge Uhren abzulesen, kann sich der Leser recht gut orientieren.

In der Erklärungsgrafik werden eine Speichengrafik und ein Balkendiagramm für die gleichen Ausgangsdaten gezeigt. Es gibt beide Male 140 Einträge. Sie sind in Siebner-Gruppen gebündelt. Die Kreislinie der Speichengrafik und die Basislinie des Balkendiagrammes haben die gleiche Länge; die Abstände der einzelnen Speichen auf der Kreislinie sind gleich gross wie die Abstände der einzelnen Balken auf der Basislinie.

Die Grafik zeigt, wie wenig Platz die Speichengrafik im Vergleich zum Balkendiagramm benötigt. Ausserdem erscheinen die Gruppen besser gebündelt; es wird schneller erkannt, wie sich die Einträge einer Gruppe zueinander verhalten. Und man sieht fast sofort, dass es etwa fünf Einträge mit sehr langen Speichen gibt.

Dadurch, dass alle Speichen gegen das Zentrum laufen, kommt es zu Überlagerungen, wenn es viele Speichen mit voller Länge gibt. Speichengrafiken eignen sich deshalb besonders für Verteilungen mit vielen kleinen bis mittleren und vereinzelter grösseren Werten.

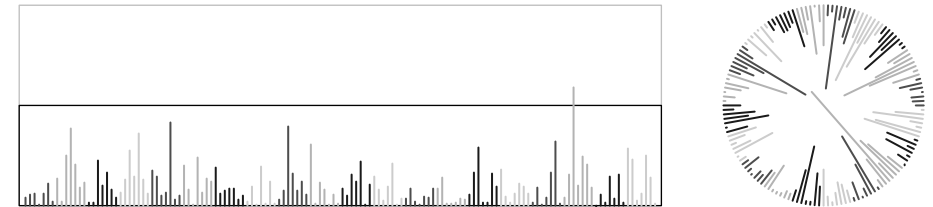
Balkendiagramme
in Kreisform

Platzsparend

► Tafel 1

Für viele kleine
bis mittlere
und vereinzelter
grössere Werte

Tafel 1: Platzbedarf einer Speichengrafik im Vergleich zum Balkendiagramm



B.3 Beispiel 1: Überblick über Fallzahlen in 445 Patientenkategorien

Einleitung

Wie kann man die Struktur des Patientenspektrums eines Krankenhauses oder einer geografischen Region grafisch darstellen?² – Mit DRG³-basierten Auswertungen können viele Details erkannt werden. Aber manchmal ist es schwierig, einen Überblick zu erhalten. Neben der grossen Anzahl von vielen hundert bis über tausend DRGs machen einem dabei auch die sehr unterschiedlichen Fallzahlen pro DRG zu schaffen: Es gibt ein paar wenige hochfrequentierte DRGs und viele schwach belegte DRGs.

² Das folgende Beispiel stammt aus Fischer [Grafiken zur PCS-Beurteilung, 2008]: 61.

³ DRG = Diagnosis Related Groups. DRG-Systeme sind Patientenkategorisierungssysteme, mit denen das Patientenspektrum stationärer Behandlungen in Akut-Krankenhäusern beschrieben werden können. In ihnen sind mehrere hundert bis über tausend Patientenkategorien definiert. – Vgl. Fetter et al. [DRGs, 1991]; Fischer [PCS, 1997]. Fischer [DRG-Familie, 2008].

Für die folgende Musterauswertung waren 20 436 Datensätze aus acht Schweizer Kinderkrankenhäusern oder -abteilungen aus dem Jahr 2005 verfügbar. Sie belegten 445 DRGs. Die Fallzahlen bewegten sich zwischen 1 und 1272 Fällen pro DRG.

Anstelle eines konventionellen Balkendiagrammes wurde ein Balkendiagramm von einer Kreislinie aus gezeichnet. Die Bezeichnungen von DRGs und Hauptkategorien wurden – hierarchisch – aussen an der Kreislinie eingetragen. – Die Grafik wurde mit dem frei verfügbaren Statistikpaket «R» erstellt.⁴

Die Grafik zeigt die Häufigkeiten pro Patientenkategorie «APDRG» (All Patient Diagnosis Related Groups). Die APDRGs sind geordnet nach Hauptkategorien. Rechts aussen beginnen die Einträge mit den APDRGs der Hauptkategorie 01 (Nerven). Die Fälle in der chirurgischen Subkategorie («01.Nerv.C») sind hellblau gefärbt, jene in der medizinischen Subkategorie («01.Nerv.M») dunkelblau. Hier findet sich auch gleich eine der sehr stark belegten APDRGs. Ihr Balken reicht zwei Kreise über die Mitte. Dieser DRG sind also mehr als 900 Behandlungsfälle zugeordnet worden. Man muss leider sehr genau hinschauen, um ihre Kurzbezeichnung ablesen zu können: Es ist APDRG 762 «Gehirnerschütterung, intrakranielle Verletzung mit Koma < 1 Std. oder ohne Koma, Alter < 18». – Am meisten Fälle gab es in APDRG 777 «Ösophagitis, gastrointestinale und verschiedene Störungen des Verdauungstraktes, Alter < 18, ohne KK», nämlich über 1200. (Deren Balken beginnt links oben und geht schräg nach rechts unten.) – Das Bündel der auffallenden, roten Balken, die von unten her kommen, stammt von den vielen Behandlungen bei Neugeborenen. – Insgesamt aber gibt es sehr viele APDRGs, die nicht einmal den ersten Kreis erreichen: Sie waren mit weniger als 100 Fällen belegt.

Die längsten Speichen kreuzen sich in der Mitte der Grafik. Damit die Grafik lesbar bleibt, darf es nicht zu viele davon geben. Die Skala der Grafik muss entsprechend dimensioniert werden.

Speichengrafiken eignen sich speziell für Daten mit vielen kleinen und mittleren sowie vereinzelt grossen Werten.

Die Beschriftungen der vielen DRGs sind sehr klein geworden. Trotzdem ist es praktisch, diese nützliche Information in «Augenreichweite» zu haben. (Nötigenfalls können die Anschriften mit einer Lupe oder mit der Zoomfunktion des Dokumentenanzeigeprogramms vergrößert werden.)

Speichengrafiken ermöglichen es, Fallzahlen von etwa 500 DRGs in einer gut lesbaren und kompakten Weise auf weniger als einer einzigen Seite darzustellen.

Daten

Grafik

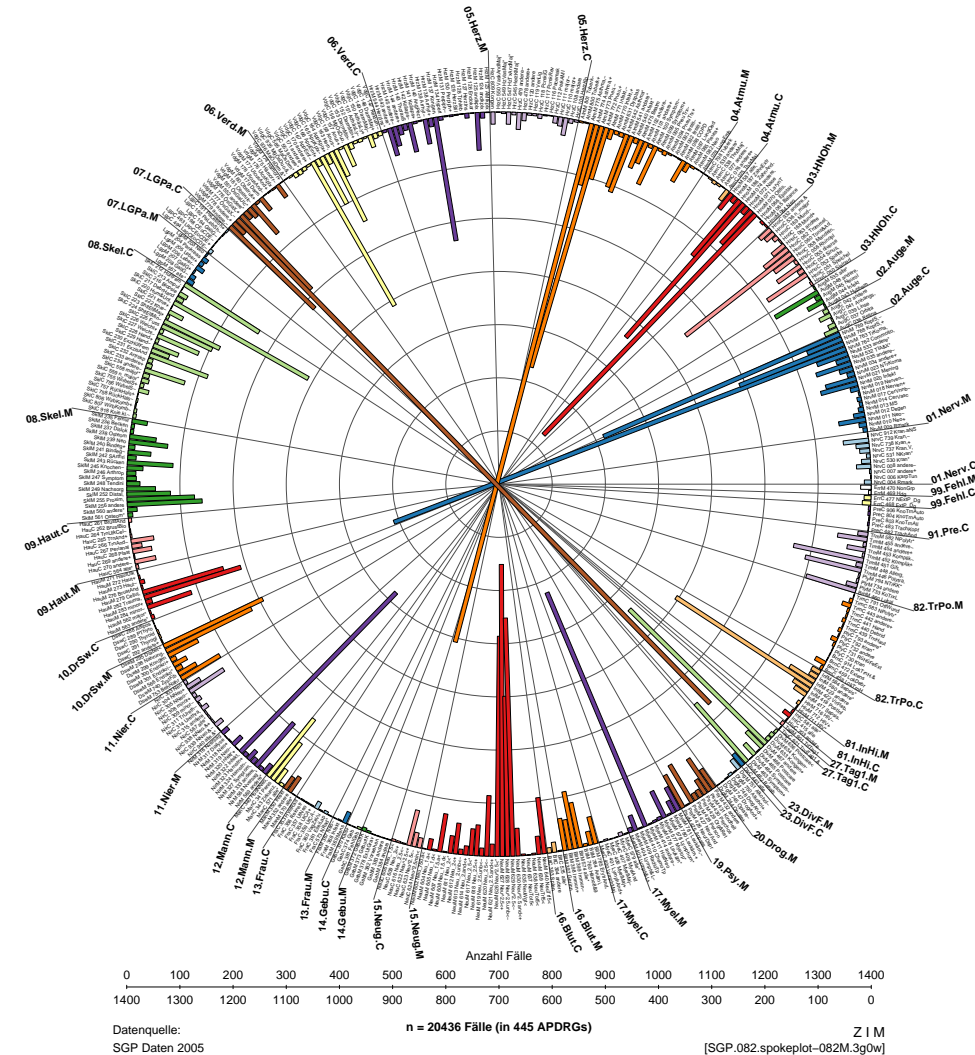
Resultat

► Tafel 2

Zur Diskussion

Konklusion

Tafel 2: SGP 2005: Anzahl Fälle in allen belegten APDRGs



⁴ <http://www.r-project.org/>; Dalgaard [R, 2002].

B.4 Beispiel 2: Preisunterschiede in 546 Patientenkategorien

Im folgenden Beispiel⁵ werden Preisunterschiede für die britischen Patientenkategorien «HRG» (Healthcare Resource Groups)⁶ gezeigt.

Für elektive und nicht- elektive Fälle werden unterschiedliche HRG-Pauschalen vergütet, die sich bei vielen HRGs deutlich unterscheiden. Die nicht- elektiven HRG-Pauschalen umfassen notfallmässige Hospitalisationen, Geburten, Neugeborene und Verlegungen.

Für jede der über 500 HRGs der Version 3.5 lagen zwei Vergütungspreise (HRG-Pauschalen) vor.⁷

Die Unterschiede zwischen elektiven und nicht- elektiven HRG-Pauschalen wurden in einer Speichengrafik visualisiert. Die Skala von der Kreislinie bis zum Kreismittelpunkt bezieht sich auf die in englischen Pfunden angegebenen HRG-3.5-Pauschalen. Die HRGs sind der Kreislinie entlang eingetragen (ohne Beschriftung). Die mit Grossbuchstaben beschrifteten Sektoren beziehen sich auf die HRG-Hauptkategorien.

Jeder Strich – ob blau oder rot – zeigt einen Unterschied zwischen der elektiven und der nicht- elektiven HRG-Pauschalen an. Identische Pauschalen sind mit grünen Punkten dargestellt. Es ist erstaunlich, dass sich die Tarife der meisten HRGs unterscheiden und dass die Unterschiede zum Teil recht gross sind. Aber es ist nicht so, dass die HRG-Pauschalen für elektive Behandlungen immer niedriger sind (blaue Striche). Dort, wo sie höher sind als die Pauschalen für nicht- elektive Behandlungen, sind rote Striche eingetragen. (Rot deshalb, weil es nicht der Erwartung entspricht, dass elektive Behandlungen teurer sind als Notfallbehandlungen.)

Das Speichendiagramm zeigt sowohl die positiven wie auch die negativen Differenzen gut.

Dadurch, dass die roten, negativen Differenzen mit etwas breiteren Linien markiert sind, wird der farbbezogene Gesamteindruck leicht verzerrt.

Von den absoluten Werten lässt sich die grobe Position zwar ablesen. Zum Beispiel zeigt der äussere Beginn der blauen Linien die elektiven HRG-Pauschalen an. Sich aber ein differenziertes Bild von der Verteilung der elektiven oder der nicht- elektiven Pauschalen machen zu wollen, ist anhand dieser Grafik schwierig. Dazu wäre es nötig, die beiden Pauschalen in einzelnen Grafiken abzubilden.

Einleitung

► Tafel 3

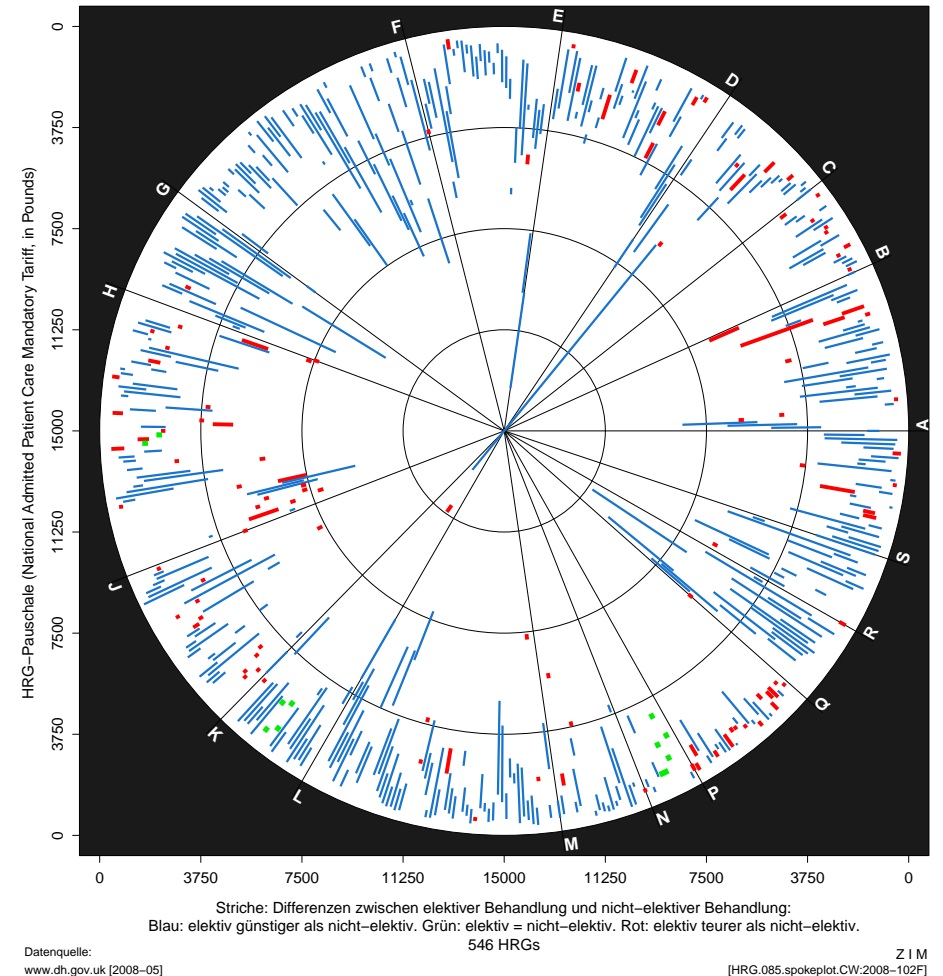
Daten

Grafik

Resultat

Zur Diskussion

Tafel 3: HRG 3.5: Unterschiede der HRG-Pauschalen 2008 für elektive und nicht- elektive Behandlungen



Quelle: Fischer [Notfallvergütung im Krankenhaus, 2009]: 95.

⁵ Das Beispiel stammt aus Fischer [Notfallvergütung im Krankenhaus, 2009]: 95.

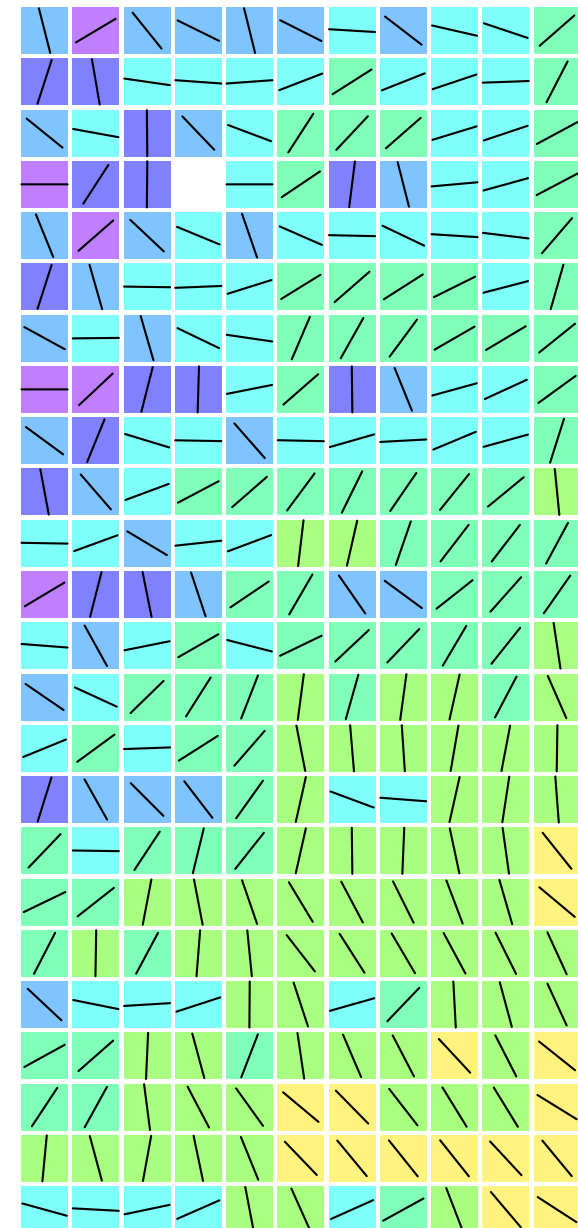
⁶ <http://www.ic.nhs.uk/our-services/standards-and-classifications/casemix>; Fischer [Notfallvergütung im Krankenhaus, 2009]: 81 ff.

⁷ DH-UK [PbR 2008/09, 2007].

C Streuungsfächerkarten

Linien-Aussaat (2009)

Gestapelte Streuungsfächerkarten ohne Streuungsfächer, zu Prämien aus mehreren Jahren. ↑ [Tafel 9 \(S. 41\)](#)

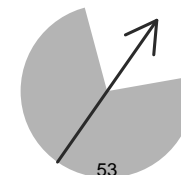


C.1 Motivation

Strömungsbilder	In einem Weltatlas sah ich Strömungsbilder von Winden. Ich war fasziniert von den Bewegungen, die hier sichtbar gemacht wurden. Spannend!, dachte ich und überlegte mir: Was wird damit eigentlich abgebildet? Die Windrichtung in Bezug auf die Himmelsrichtung und die Windgeschwindigkeit. Was das Ganze interessant und praktisch verwendbar macht, ist, dass die beiden Variablen durch die Angabe der geografischen Position verortet sind. Das alles zusammen erscheint in einer attraktiven Darstellung. Wie wäre es, begann ich zu überlegen, wenn man das Prinzip auf andere Daten anwenden würde?
x- und y-Skala	Für die x- und y-Dimensionen der Zeichenfläche wären nicht nur metrische, sondern auch ordinale Skalen geeignet wie etwa ordinal geordnete Kategorien (z. B. Schweregrade einer Erkrankung) oder nach Klassen unterteilte Intervallskalen (z. B. Alterstufen).
Drehbare Pfeile ↑ Vgl. Tafel ?? (S. ??)	Wenn ich nun einen Pfeil in jedes der so entstehenden Felder zeichnen würde, hätte ich schon ein abstrahiertes «Strömungsbild». Was aber soll die Drehrichtung des Pfeiles und was dessen Länge darstellen?
Länge	Die Länge kann zur Abbildung der Anzahl Messwerte im betreffenden Feld verwendet werden. Kleinere Pfeile zeigen dann auf kleinere und weniger gewichtige Gruppen. Die Grösse wird damit zu einer Relevanzanzeige.
Drehrichtung	Für die Drehrichtung verbleibt der Messwert selbst. Je nach der Wahl von x- und y-Skala kann es sein, dass ein Feld mehrere Messungen betrifft. In solchen Fällen könnte der Mittelwert oder der Median abgebildet werden. Aber wenn schon Mittelwert oder Median dargestellt werden, könnte dann nicht auch noch ein Streuungsmass angezeigt werden? Aus dieser Überlegung heraus schnitt ich den Hintergrund so aus, dass damit der Interquartilsbereich sichtbar wurde.
↑ Median: S. 99	
► Tafel 4	
Geburt des Fächers	In weiteren Versuchen mit diesem Grafiksymbol zeigte sich, dass sich die Pfeilspitze und der Interquartils-Ausschnitt ungünstig konkurrenzieren, denn sie sind gegenläufig. Ausserdem ist eine Pfeilspitze ein vielleicht unnötiger Schnörkel; ¹ man müsste nur die Ausrichtung anders klarstellen

¹ Tufte unterscheidet zwischen Druckfarbe, die zur Wiedergabe der Daten unverzichtbar nötig ist, und solcher, die sonst für die Grafik verwendet wird. Das Verhältnis dieser beiden Mengen sollte möglichst gross sein, d. h. es sollte weggelassen werden, worauf ohne Datenverlust verzichtet werden kann. – Tufte [Display, 2001]: 93 ff.

Tafel 4:
Erster Versuch zu
einem grafischen
Symbol für ein
Strömungsbild



können. Lange suchen musste ich nicht mehr: Die Öffnung des Ausschnittes zeigte ja bereits die Richtung. Öffnung des Ausschnittes und Schaft des Pfeiles aber ergeben – zusammengekommen – eine neue Figur: Der Fächer war geboren. Ich färbte ihn schwarz. Damit hatte ich die Buntfarben frei für eine andere Variable.

Nun war es nur noch ein kleiner Schritt, die vom Boxplot her bekannten weiteren Informationen hinzuzufügen: Links und rechts vom schwarzen Streuungsfächer positionierte ich zwei weisse Federn. Sie kennzeichnen das 5 %- und das 95 %-Perzentil, umfassen also 90 % der Werte.²

↑ Boxplot: S. 99

C.2 Beschreibung: Fächer

Die «Streuungsfächerkarte»³ ist eine Grafik, deren Gestaltung von meteorologischen Windströmungskarten inspiriert wurde. Anstelle eines Pfeiles, welcher die Windrichtung anzeigt, wird ein gerichtetes Fächersymbol verwendet. Es zeigt Anzahl (Länge der vorderen, weissen Mittellinie), Median (Symbolausrichtung) und Streuung (schwarzer Fächer und weisse Federn) der Messwerte. Es kann anstelle von Boxplots verwendet werden.

► Tafel 5

Auf einer zweidimensionalen Fläche können die Verteilungskennzahlen für Untergruppen einer Stichprobe, welche nach zwei Variablen klassiert ist, dargestellt und direkt verglichen werden.

Als Skala wird eine Kreislinie benutzt. Die Skala beginnt ganz links, z. B. mit Null. Die weiteren Werte werden im Uhrzeigersinn aufgetragen.

Skala auf Kreislinie

Mit dem Fächer-Symbol werden im Einzelnen abgebildet:

Abbildung von
Anzahl, Median
und Quartilen

- Ausrichtung des Symbols: Median.
- Breite des Fächers: Quartilsabstand: Erstes bis drittes Quartil (25 %- bis 75 %-Perzentil).
- Äussere weisse Linien («Federn»): 5 %- bis 95 %-Perzentil.
- Länge der vorderen Mittellinie (weisse Linie im schwarzen Fächer): Anzahl Messwerte.
- Dicke der Mittellinie (sichtbar als Griff des Fächers): Proportional zum Logarithmus der Anzahl Messwerte.
- Hintergrundfarbe: Median (= Ausrichtung des Symbols).
- Zahlen am unteren Rand: Median sowie Anzahl Messungen.

² Je nach Bedarf kann diese Grenze verändert werden. – Im Boxplot nach Tukey werden die «whiskers» beim $\langle 3. \text{Quartil} + k \times \text{IQR} \rangle$ bzw. beim $\langle 1. \text{Quartil} - k \rangle \times \text{IQR}$ gesetzt (IQR = Interquartilsabstand). Tukey schlug $k = 3.0$ vor; heute wird oft $k = 1.5$ verwendet. – Tukey [EDA, 1977], zitiert in: Nagel et al. [Grafische Datenanalyse, 1996]: 35 ff.

³ Englisch: «fan chart».

↑ Beispiel:
Tafel 7 (S. 37)

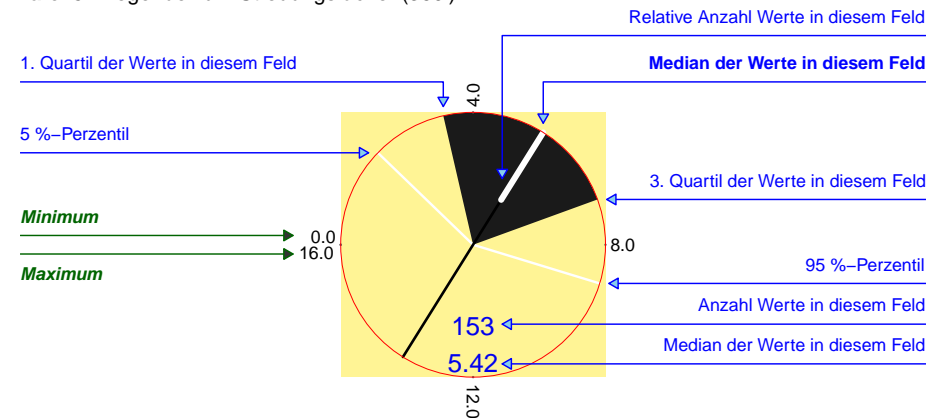
Bei der Interpretation einer Streuungsfächerkarte wird nach Abhängigkeiten der Ausrichtungen von einer oder von beiden klassierenden Variablen gesucht. Dazu lässt man den Blick am Besten zeilenweise, dann spaltenweise und dann noch diagonal durch das Bild laufen und kontrolliert, wie sich dabei die Ausrichtungen der Fächer drehen.

- Dreht sich die Ausrichtung kontinuierlich im Uhrzeigersinn, d. h. zu grösseren Messwerten hin, so korreliert die Abhängigkeit der Messwerte positiv mit der klassierenden Variablen, die in der Bewegungsrichtung des Blickes aufgetragen ist.
- Dreht sich die Ausrichtung kontinuierlich im Gegenuhrzeigersinn, d. h. zu kleineren Messwerten hin, so korreliert die Abhängigkeit der Messwerte negativ mit der klassierenden Variablen, die in der Bewegungsrichtung des Blickes aufgetragen ist.

Zusätzlich kann beachtet werden:

- Je geschlossener der Fächer ist (d. h.: je kleiner die Streuung ist), desto konzentrierter liegen die Messwerte um den Median der mit dem Fächersymbol abgebildeten Untergruppe.
- Je länger der weisse vordere Teil der Mittellinie ist (d. h.: je grösser die Anzahl der Messwerte ist), desto verlässlicher sind Median und Streuung der Messwerte in der mit dem Fächersymbol abgebildeten Untergruppe.

Tafel 5: Legende zum Streuungsfächer (360°)



C.3 Beispiel 1: Fähigkeitseinschränkungen und Pflege

In der neurologischen Rehabilitation war unter anderem zu prüfen, ob sich der Pflegeaufwand aufgrund der Kenntnisse der Fähigkeitseinschränkungen schätzen liesse. Falls dies der Fall wäre, könnte man die Messung von Fähigkeitseinschränkungen auch für ein Tarifierungssystem benutzen. Ein solches System wurde im Rahmen des TAR-Projektes entwickelt.⁴

In der Grafik wird die Abhängigkeit des Pflegeaufwandes von den Fähigkeitseinschränkungen der Patienten gezeigt. Auf der x-Achse sind die FIM-Werte zu den kognitiven Fähigkeitseinschränkungen, auf der y-Achse die FIM-Werte zu den praktisch-motorischen Einschränkungen gruppiert.⁵ In den Feldern sind neben dem Streuungsfächer die Anzahl der Patientenwochen, der Medianwert des Pflegeaufwandes pro Tag und – rot – die Anzahl der Wochen mit einem durchschnittlichen Tagesaufwand von mehr als acht Pflegestunden (gemessen mit dem Instrument LEP⁶) eingetragen. Die Farbe der Felder korrespondiert mit dem Median des Pflegeaufwandes.

Links unten, ausserhalb des Grafikfeldes, ist als Legende ein kleiner Fächer auf grauem Hintergrund aufgezeichnet. Darauf ist die Skala ersichtlich sowie die Streuung der gesamten Stichprobe.

Als Erstes fällt die Abnahme des Pflegeaufwandes auf den meisten Zeilen von rechts nach links auf: Von gut fünf und mehr Stunden zu weniger als einer Stunde. Das heisst, mit zunehmenden praktisch-motorischen Fähigkeiten ist weniger Pflege nötig. Das entspricht der Erwartung: Je mehr der Patient selber tun kann, desto weniger muss er gepflegt werden.

Bezüglich der kognitiven Fähigkeiten ist die Sachlage komplexer. Es fallen die drei violetten Felder mit über sechs Stunden auf. Sie liegen alle am Rand, d.h. diese Patienten sind entweder praktisch-motorisch oder kognitiv völlig unselbständig. In allen diesen Feldern ist die Streuung sehr gross.

Die insgesamt unregelmässige Verteilung weist darauf hin, dass eine Schätzung des Pflegeaufwandes mittels linearer Regression vermutlich zu grob ausfällt und eine allenfalls darauf aufbauende Vergütung deshalb als inadäquat empfunden werden könnte.

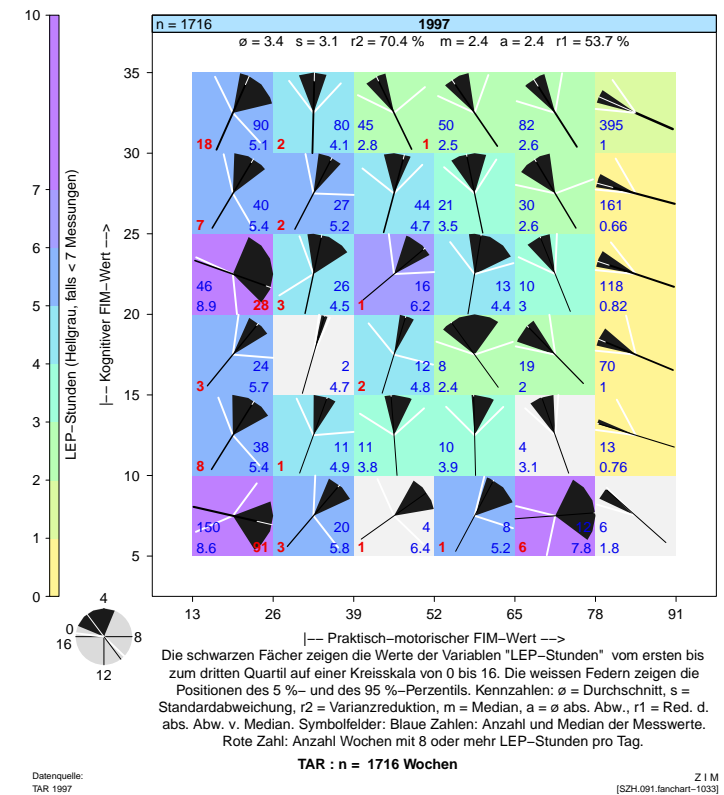
Einleitung

Grafik
► Tafel 6

Resultate

Zur Diskussion

Tafel 6:
Median und Quartile der LEP-Stunden pro Tag, nach motorischem und kognitivem FIM (TAR97)



⁴ TAR = Leistungsbedarfsbezogenes Tarificationssystem für Rehabilitationskliniken. – Fischer et al. [TAR und Reha-PCS, 2006].

⁵ FIM = Functional Independence Measure. Mit dem FIM-Instrument werden 13 Items zur praktisch-motorischen und 5 Items zu kognitiven Fähigkeitseinschränkungen von Patienten auf einer einheitlichen 7-stufigen Skala beurteilt. Die praktisch-motorische Skala reicht somit von 13 bis 91, die kognitive von 5 bis 35 FIM-Punkten. – Granger et al. [FIM, 1995]. UDSmr [FIM™ instrument, 1997].

⁶ LEP = Leistungserfassung in der Pflege. Mit LEP werden die erbrachten Leistungen der Pflegefachpersonen nach einer Tätigkeitsliste mit gut 80 Einträgen täglich erhoben, mit Normzeiten gewichtet und pro Patient aufaddiert. – Vgl. Brügger et al. [LEP-Methode 2.0, 2002]; Fischer [DRG+Pflege, 2002]: 141 ff + 247 ff.

C.4 Beispiel 2: Fähigkeitseinschränkungen und Pflege im Zeitverlauf

In der folgenden Grafik mit Daten aus einer anderen Erhebung zur Neuro-Rehabilitation können die Abhängigkeiten des Pflegeaufwandes von den kognitiven und praktisch-motorischen Fähigkeitseinschränkungen aus vier Jahren miteinander verglichen werden.

Zur Unterteilung in Gruppen wurde die Technik der Fachwerkgrafiken benutzt. Damit werden viele Einzelgrafiken («Grafikfelder») matrixförmig aneinander gefügt und gleichzeitig auf einer einzigen Seite dargestellt.⁷

Im Überschriftsbalken ist links von der Anschrift jedes Grafikfeldes die Anzahl Messwerte eintragen. Der hellblaue Balken, der von rechts her kommt, zeigt den Anteil der in diesem Grafikfeld abgebildeten Messwerte im Verhältnis zu allen Messwerten. Die Länge des Balkens kann als Indikator für die Relevanz der Auswertungen im jeweiligen Grafikfeld verwendet werden.

Die hellgrauen und die leeren Felder sowie die eher diagonal liegenden farbigen Felder auf den Streuungsfächerkarten zeigen, dass es nicht viele Behandlungen gab, wo die kognitive Beeinträchtigung und die praktisch-motorische Beeinträchtigung extrem gegensätzlich waren.

Auf den Streuungsfächerkarten wechselt die Farbe von rechts nach links von gelb zu blau und vereinzelt sogar bis violett. Dies weist auf eine deutliche Zunahme der LEP-Stunden pro Tag bei zunehmenden praktisch-motorischen Fähigkeitseinschränkungen hin. Dem entsprechend dreht sich die Richtung des Fächers auf den meisten Zeilen kontinuierlich, wenn einzelne «Grafikzeilen» mit dem Auge durchwandert werden. Vereinzelt bewegt sich der Fächer auch mal in die Gegenrichtung.

Von oben nach unten kann man weniger Unterschiede feststellen. Die Fächer zeigen in den meisten Spalten in eine jeweils ähnliche Richtung. Wenn sie sich drehen, tun sie das des Öftern nicht kontinuierlich mit dem gleichen Drehsinn. Das weist darauf hin, dass der Pflegeaufwand nicht direkt vom Ausmass der kognitiven Beeinträchtigungen gemessen in FIM-Punkten abhängig ist.

Für den Überblick ist die Grafik gut geeignet. Für eine detailliertere Analyse sollte sie mit Zahlen versehen und auf einem grösseren Blatt ausgedruckt werden.

Die schwarzen Kleckse der übermässig breiten Fächer auf farbigem Grund regen an, sich die Daten der betreffenden Patienten näher anzuschauen. War der Pflegeaufwand hier wirklich so unterschiedlich? Warum? Oder liegt ein Datenfehler vor?

Grafik
► Tafel 7

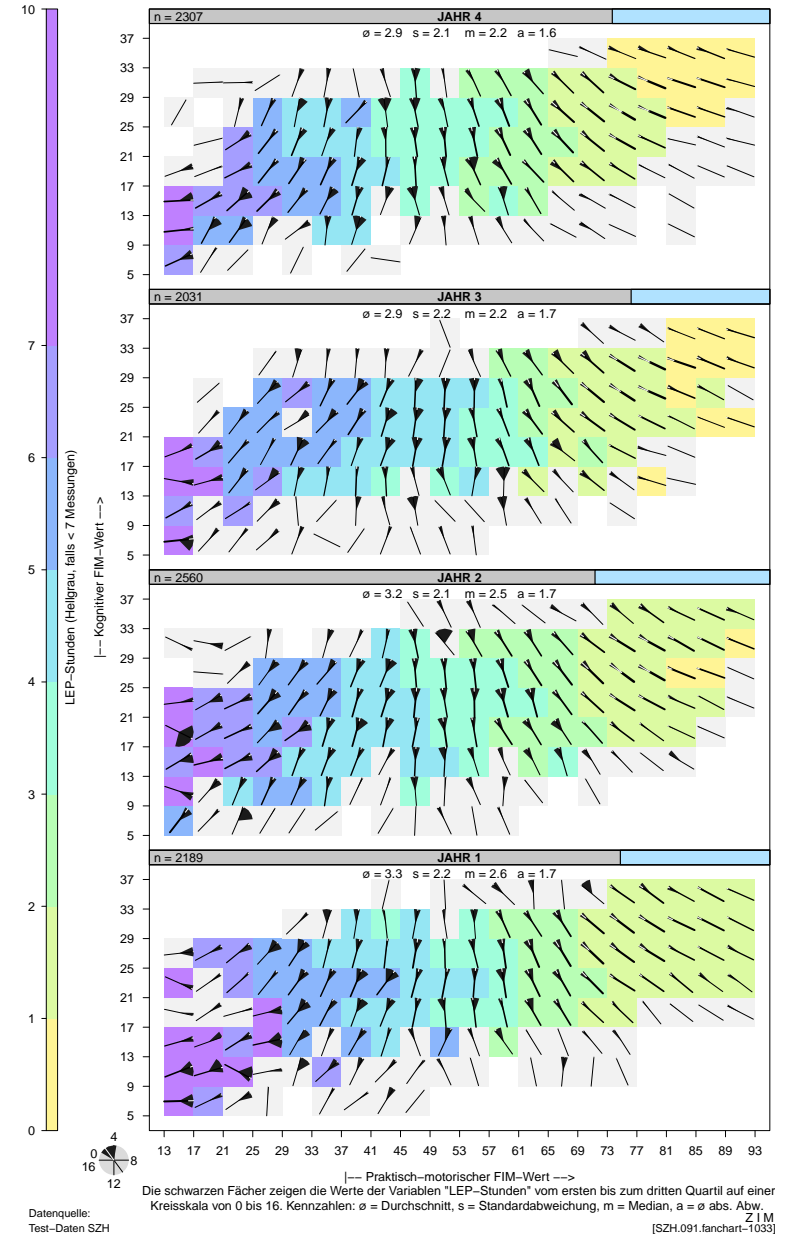
Fachwerkgrafik

Relevanzbalken

Resultate

Zur Diskussion

Tafel 7: Median und Quartile der LEP-Stunden pro Tag, nach motorischem und kognitivem FIM (Test-Daten)



⁷ Englisch: Trellis displays. – Vgl. Becker et al. [Trellis, 1996]; Becker/Cleveland [Trellis/Man, 1996]; Cleveland [Visualizing, 1993] und <http://cm.bell-labs.com/cm/ms/departments/sia/project/trellis/>. – Für Fachwerkgrafiken gibt es im R-Projekt das Paket «lattice».

C.5 Beispiel 3: Mathematikprüfungen und sozioökonomischer Status

Das für das folgende Beispiel verwendete Datenset stammt aus dem R-Paket «nlme».⁸ Es beinhaltet Resultate zu 7185 Mathematikprüfungen. Die StudentInnen sind kategorisiert nach Geschlecht und Zugehörigkeit zu einer ethnischen Minorität.⁹

In der Grafik wurden die Prüfungsergebnisse in Abhängigkeit vom sozioökonomischen Status der StudentInnen (x-Achse) und vom durchschnittlichen sozioökonomischen Status der Schule (y-Achse) abgebildet. In den vier Hauptgrafikfeldern werden StudentInnen nach ihrem Geschlecht und nach der Zugehörigkeit zu einer Minorität unterschieden.

Als Erstes fallen die weit geöffneten Fächer auf, d. h. die grossen Streuungen der Prüfungsergebnisse in praktisch allen Untergruppen. Der Vergleich der vier Hauptgrafikfelder zeigt grundsätzlich bessere Prüfungsergebnisse bei StudentInnen, die nicht einer Minorität angehören, sowie leicht bessere Resultate der Männer im Vergleich zu den Frauen.

Innerhalb der Felder ist festzustellen, dass sowohl der sozioökonomische Status der StudentInnen als auch – etwas geringer – der durchschnittliche sozioökonomische Status der Schule positiv mit den Prüfungsergebnissen korrelieren. – Bei der Detailbetrachtung zeigt sich, dass es auch Männer und Frauen aus Minoritäten gibt, die sehr gute Resultate haben können. (Vgl. innerhalb der rechten Grafikfelder die blau gefärbten Untergruppenfelder rechts oben mit bis weit nach unten geöffneten Fächern.)

Es stellt sich die Frage, ob zur Abbildung des durchschnittlichen sozioökonomischen Status der Schule die gleiche Skaleneinteilung verwendet werden sollte wie für die Abbildung des sozioökonomischen Status der StudentInnen. Da der Status der Schule als Durchschnitt aus den Werten der StudentInnen berechnet wird, streut er weniger: Der Interquartilsbereich liegt zwischen -0.32 und 0.33 beim durchschnittlichen Status der Schulen und zwischen -0.54 und 0.60 beim Status der StudentInnen. Die beiden Variablen weisen also eine unterschiedliche Verteilung auf. Deshalb können auch die Skaleneinteilungen unterschiedlich gewählt werden.

Je nach Wahl der Skaleneinteilungen verändern sich allerdings die Streuungsfächer. Die Wahl erfolgte so, dass die Bewegungen der Streuungsfächer sichtbar wurden, ohne dass die Anzahl StudentInnen in den Untergruppen zu klein wurde.

Diese Grafik zeigt anschaulich, wie der mittlere Wert z. T. einer grossen Haupttendenz folgt, obwohl die Werte in den einzelnen Untergruppen stark streuen und man deshalb über eine mögliche Korrelation ins Zweifeln geraten könnte. (Besonders gut ist dies in den unteren beiden Feldern sichtbar.)

Daten

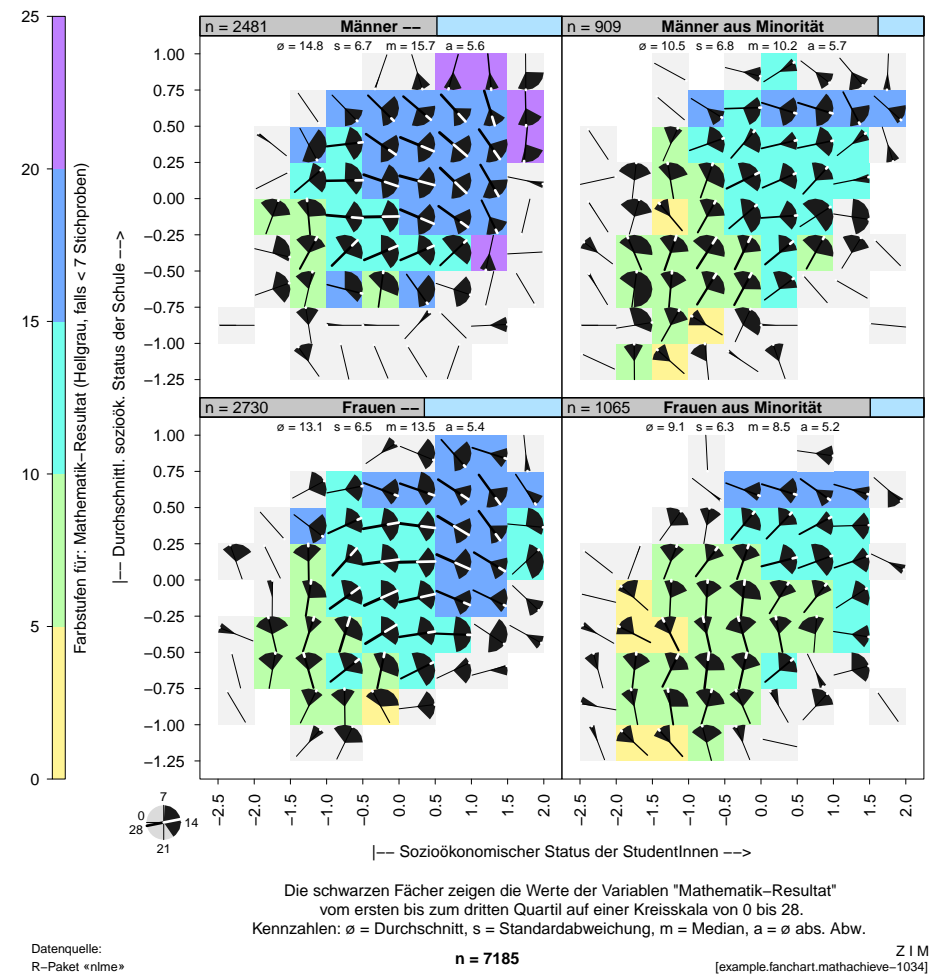
Grafik
► Tafel 8

Resultate

Zur Diskussion

Konklusion

Tafel 8: 7185 Mathematikprüfungen: Resultate nach Geschlecht und Zugehörigkeit zu einer Minorität



⁸ <http://www.r-project.org/>: «Linear and Nonlinear Mixed Effects» von Pinheiro J et al.

⁹ Im amerikanischen Original: «member of a minority racial group».

C.6 Beispiel 4: Krankenkassenprämien

In der Schweiz geben die Krankenversicherer («Krankenkassen») jedes Jahr im Herbst die Krankenkassenprämien für das Folgejahr bekannt.¹⁰

Die Krankenpflegeversicherung ist seit 1996 für alle Bewohner des Landes obligatorisch.¹¹ Jeder versichert sich selbst bei einer Krankenkasse seiner Wahl. Die obligatorisch zu versichernden Leistungen beinhalten die ambulante und stationäre medizinische Behandlung im Wohnkanton.

Die Krankenversicherer befinden sich in einer Konkurrenzsituation und sind frei in der Festlegung der Prämien, solange sie eine gesunde wirtschaftliche Lage ihrer Kasse sicherstellen. Die Prämien für Erwachsene dürfen nicht nach Altersklassen und Geschlecht, wohl aber nach Kanton und innerhalb mancher Kantone nach bis zu drei vom Bund vordefinierten «Prämienregionen» differenziert werden. Es ist erlaubt, für «junge Erwachsene» und für «Kinder» Rabatte zu gewähren.

In der Grafik sieht man die Streuung der Prämien für grundversicherte Erwachsene, welche die in den einzelnen Kantonen tätigen Krankenversicherer pro Region anbieten.

Die Anordnung der Felder ist «pseudogeografisch», d. h. die Kantone und deren Prämienregionen sind in ähnlichen Gegenden zu finden wie auf einer Schweizer Karte. Zum Beispiel ist Genf («GE») ganz links unten positioniert und die beiden Tessiner Regionen («TI1» und «TI2») sind unten rechts von der Mitte zu finden.

Aus der Streuungsfächerkarte ist zunächst ersichtlich, dass die Prämien in der französischen Schweiz (eher links im Diagramm), im Tessin und in den städtischen Regionen (Basel Stadt: «BS», Zürich Region 1: «ZH1», usw.) tendenziell höher sind als in den ländlicheren Kantonen. Die Streuungen der Prämien sind erstaunlich gross. Als Ausnahme fällt Genf (ganz links unten) auf, wo die mittleren 50% der Prämienangebote praktisch gleich sind. (Der schwarze Fächer ist ganz geschlossen.)

Infolge der pseudogeografischen Anordnung haben die x- und y-Skalen hier eher nominalen als ordinalen Charakter. Deshalb macht eine zeilen- oder spaltenweise Betrachtung wenig Sinn. Aus etwas Distanz kann man aber trotzdem die höheren Prämien im Westen des Landes (d. h. eher links auf der Streuungsfächerkarte) erkennen.

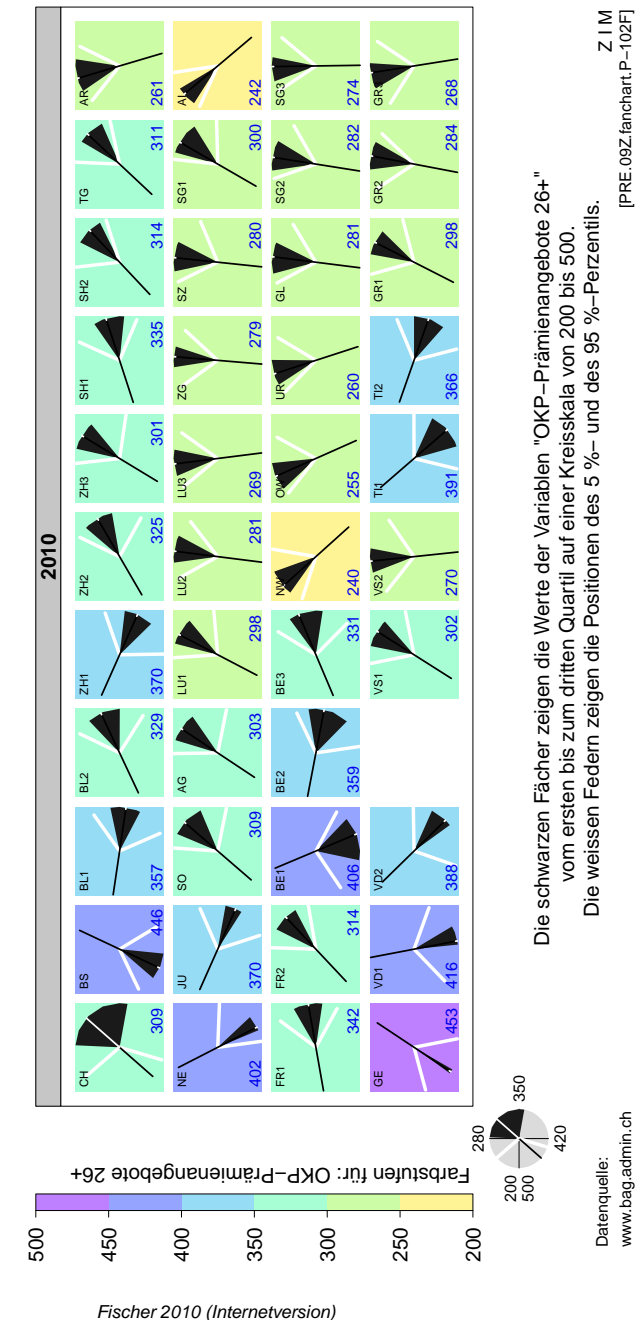
Einleitung

Grafik
► Tafel 9

Resultate

Zur Diskussion

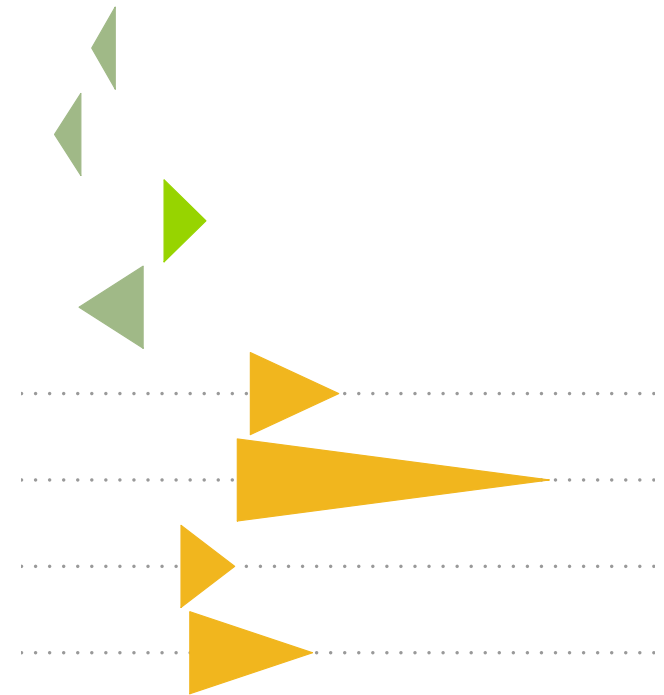
Tafel 9: CH: Krankenkassenprämienangebote



¹⁰ Vgl. <http://www.praemien.admin.ch/>.

¹¹ Art. 3 KVG.

D Differenzdiagramme



Dreiecke im Wind (2005)

Ausschnitt aus Differenzdiagramm mit Fallanteilen zur APDRG-Hauptkategorie 08. ↑ Tafel 11 (S. 49)

D.1 Motivation

Absolute und/oder relative Werte?

Bei der Darstellung verschiedener Zahlenreihen steht man oft vor der Frage, ob man die absoluten Werte, die absoluten Differenzen oder die prozentualen Differenzen anzeigen soll. Wenn es ums Vergleichen geht, sind Differenzen die Zielgrösse. Zeigt man aber nur die Differenzen, fehlen einem Bezugswerte.

Balkendiagramme?

Deshalb stellt man gerne Balken nebeneinander: Dann sieht man die absoluten Werte, und aus den unterschiedlichen Längen der Balken lassen sich die Differenzen ablesen. Ausserdem gewinnt man eine Idee von der prozentualen Differenz, wenn man die Differenz zweier Balken mit der Gesamtlänge des einen dieser Balken vergleicht.

Positive und negative Differenzen

↑ Beispiel mit Differenzbalken: Tafel 3 (S. 25)

Wenn man der Differenz eine grössere Wichtigkeit geben will, könnte man versucht sein, nur einen einzigen Balken zu verwenden und damit die Differenz abzubilden. Positive und negative Differenzen könnte man unterschiedlich färben. Bei diesem Vorgehen sieht man nun zwar die Differenzen gut; man hat aber Mühe, wenn man die absoluten Werte den Balkenenden zuordnen will, da man erst überlegen muss, zu welchem Ende welche Werte nun gehören.

Dreiecke sind ausgerichtet

Wenn man anstelle solcher Differenzbalken Dreiecke verwendet, werden auch die absoluten Werte wieder besser sichtbar: Man kann die Basen der Dreiecke oder die Spitzen verfolgen.

Mit aneinander gereihten Dreiecken benötigt man weniger bedruckte Fläche als mit nebeneinander gestellten Balken. Dies macht die Grafiken deutlich übersichtlicher.

D.2 Beschreibung: Dreiecke

Basis und Spitze

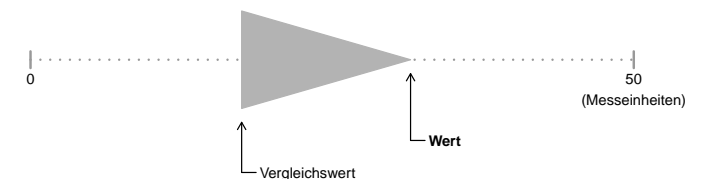
► Tafel 10

Da ein Dreieck eine Ausrichtung hat, können damit zwei unterschiedliche Werte auf einer gemeinsamen Skala angezeigt werden: Ein Wert mit der Basis der Dreiecks und ein Wert mit der Spitze des Dreiecks.

Fläche

Die Fläche, welche das Dreieck aufspannt, ist proportional zur Differenz zwischen den beiden Werten. Je mehr Farbe erscheint, desto grösser ist die Differenz.

Tafel 10:
Legende zum Symbol für Differenzdiagramme



Positive und negative Unterschiede können durch unterschiedliche Färbungen angezeigt werden. Dank den Strassenampeln gut verständlich ist z. B. die Benutzung der Farben Grün und Rot.

Farben

D.3 Beispiel 1: Vergleich von Fallanteilen in Patientenkategorien

Für die folgende Grafik lautete die Arbeitsfrage: Wie kann man die Besonderheiten des Patientenspektrums eines Krankenhauses im Vergleich zu den Patientenspektren anderer Krankenhäuser des gleichen Typs darstellen?¹

Einleitung

Als Datenbasis wurden vom Krankenhaus, das zu analysieren war, und von allen 27 Vergleichskrankenhäusern Falldatensätze mit DRG²-Codes benötigt. Die DRG-Codes wurden nach Hauptkategorien («01» = Nerven, «02» = Augen, . . .), Subkategorien («M» = medizinisch, «C» = chirurgisch) und Kostengewichtsklassen («1» bis «4») kategorisiert.

Daten
► Tafel 11

Für die Grafik wurden für jede Kategorie die Differenz zwischen dem Fallanteil des zu untersuchenden Krankenhauses und dem Durchschnitt aller Krankenhäuser in der Vergleichsgruppe berechnet. Sie wurde mit einem Dreieck abgebildet. Negative Abweichungen vom Durchschnitt wurden mit dunkleren, positive mit klaren Farben markiert. Medizinische Behandlungen wurden grün, chirurgische goldfarben gefärbt.

Grafik

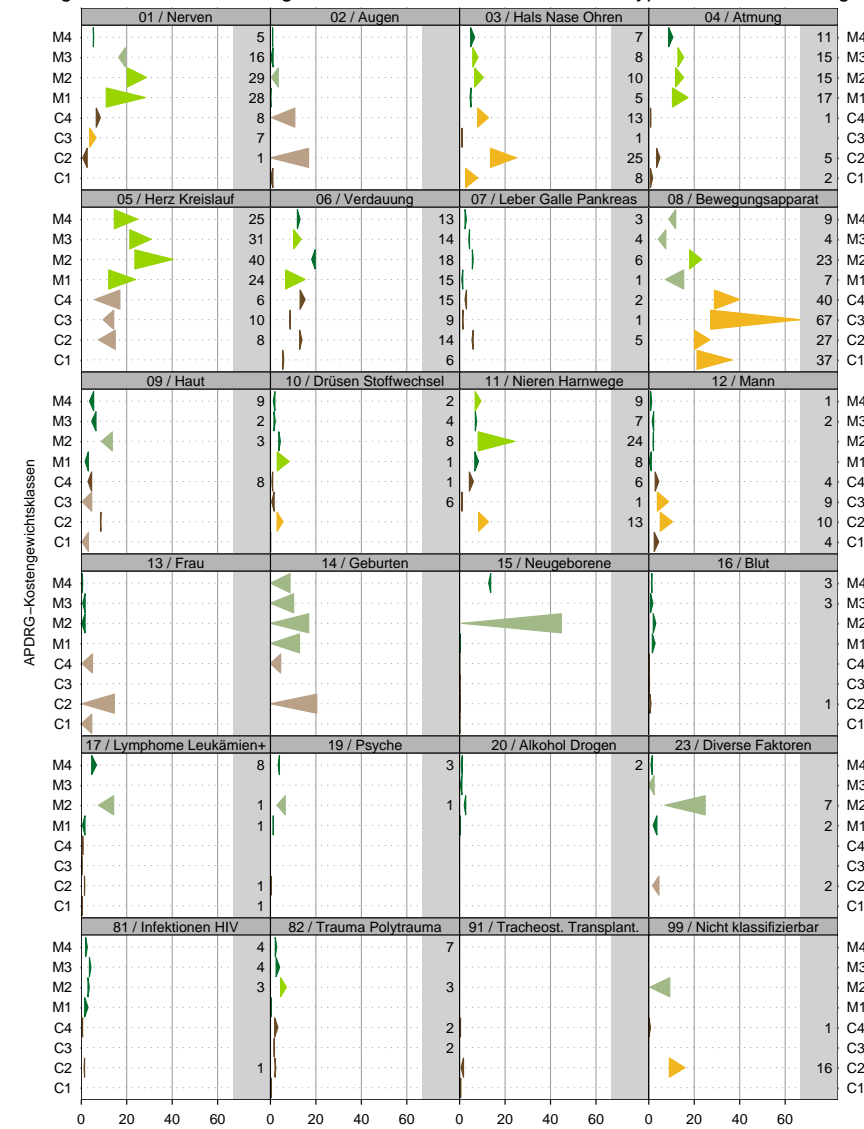
Die Grafik erlaubt es, auf den ersten Blick negative und positive Abweichungen in medizinischen und chirurgischen Kostengewichtsklassen der verschiedenen DRG-Hauptkategorien zu erkennen.

Resultate

Besonders auffällig sind die überdurchschnittlichen Fallanteile bei den medizinischen Behandlungen am Bewegungsapparat («08») sowie bei den chirurgischen Eingriffen am Herz und am Kreislaufsystem («05»). Bei den medizinischen Behandlungen an Nieren und Harnwegen («11») sind nur DRGs in der zweitgünstigsten Kostengewichtsklasse («M2») überdurchschnittlich vertreten; Behandlungen mit DRGs in den übrigen Kostengewichtsklassen entsprechen in etwa dem Durchschnitt.

Bereiche ohne Behandlungen sind an Dreiecksspitzen, die an den linken Grafikfeldrand stoßen, oder an leeren Zahlenspalten zu erkennen. Dies gilt z. B. für die Augen («02»), für Geburten («14») und Neugeborene («15»).

Tafel 11: Krankenhaus H217: Fallanteile nach APDRG-Subkategorien und Kostengewichtsklassen im Vergleich zu allen Krankenhäusern des Typs «Zentrumsversorgung»



Dreiecke und Zahlen: Fallanteile (in Promille) von H217.
Basis der Dreiecke: Durchschnitt aller 27 Spitäler des Typs

Datenquelle: Bundesamt für Statistik:
Medizinische Statistik der Krankenhäuser

K11 (Zentrumsversorgung).

Z 1 M - Test . 31
[BFSMS.042.H.diffplot.nf.CWcat-2002-0530]

Quelle: Nach Fischer [Krankenhaus-Betriebsvergleiche, 2005]: 45.

¹ Das Beispiel stammt aus: Fischer [Krankenhaus-Betriebsvergleiche, 2005]: 45.

² DRG = Diagnosis Related Groups. Mit DRG-Systemen können Patientenspektren akutstationärer Behandlungen beschrieben werden. – Vgl. Fetter et al. [DRGs, 1991]; Fischer [PCS, 1997]. Fischer [DRG-Familie, 2008].

D.4 Beispiel 2: Gerstenernten 1931 und 1932

Daten zu Gerstenernten von 1931 und 1932 in Minnesota wurden seit 1934 von amerikanischen Statistikern wiederholt ausgewertet. Cleveland machte in seinem Buch von 1993 auf eine «Datenanomalie» der Daten aus «Morris» aufmerksam. Anschliessend belegte er seine Vermutung, dass die Daten zu den beiden Jahren von «Morris» vertauscht seien u. a. auch anhand von grafischen Auswertungen.³ Interessant ist, dass Cleveland offenbar der erste war, dem die Unterschiede auffielen und der sie nicht einfach als gegeben hinnahm.

Die hier verwendeten Daten bestehen aus Ertragsmengen für die Jahre 1931 und 1932 von sechs Anbauorten (von «Waseca» bis «Rapids») und zehn Gerstensorten (von «Treb» bis «No. 475»).⁴ Der Datensatz umfasst 120 Messungen (= zwei Jahre × sechs Orte × zehn Sorten).

Zur Darstellung wählte ich ein Differenzdiagramm. Anbauorte und Gerstensorten sind geordnet nach den Erträgen von 1931. Die Erträge von 1931 werden mit der Basis der Dreiecke abgebildet, die Erträge von 1932 mit der Spitze. Die Länge der Dreiecke entspricht dem Ertragsunterschied. Verminderte Erträge sind rot, erhöhte Erträge grün eingefärbt.

Es fällt auf den ersten Blick auf, dass die Daten aus «Morris» im Unterschied allen den übrigen Anbauorten durchgehend erhöhte Erträge aufweisen. Es zeigt sich weiter, dass in «Duluth» die kleinsten Ertragsseinbusen zu verzeichnen waren. Allerdings war hier die Ernte von 1931 bereits relativ niedrig, d. h. insbesondere niedriger als in «Waseca» und in «Crookston», aber ähnlich hoch wie in «Rapids».

Um die 120 Messungen überblicken zu können, sind die bereits von Cleveland benutzten Fachwerkgrafiken⁵ äusserst nützlich.

Eine andere Möglichkeit, die beiden Ernten in einem einzigen Grafikfeld darzustellen, ist die Verwendung von graphischen Zeichen wie «o» und «+». Diese Darstellungsart benutzte Cleveland. Wenn man genau hinschaut, ist auch in solchen Darstellungen ersichtlich, dass die Zeichen für «Morris» vertauscht sind. Im Differenzdiagramm jedoch wird die Besonderheit von «Morris» durch die unterschiedliche Einfärbung und Ausrichtung der Dreiecke bei Abnahmen und Zunahmen unmittelbar augenfällig.

Einleitung

► Tafel 12

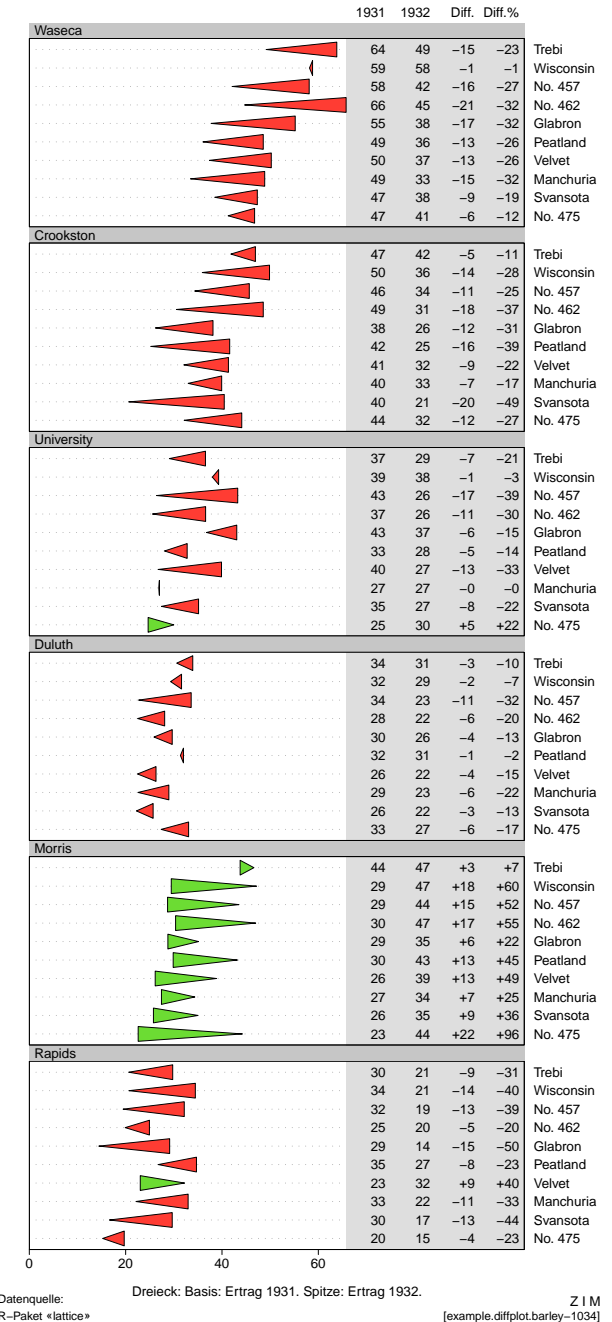
Daten

Resultate

Zur Diskussion

↑ Tafel 15 (S. 55),
2. Spalte

Tafel 12:
Gerstenernten in
Minnesota 1931
und 1932 (Diffe-
renzdiagramm)



³ Vgl. Cleveland [Visualizing, 1993]: 328 ff., insbesondere Tafel 6.20 (S. 330) und 338 f.

⁴ Es wurden die Daten verwendet, die dem R-Paket «lattice» beigegeben sind. Der angegebene Ertrag ist der durchschnittliche Ertrag von jeweils drei zufällig ausgewählten Anbaufeldern je Anbauort. Gemessen wurde er in Scheffel (US-«bushels» = 35 Liter × ca. 30 kg ≈ 1050 kg) pro Morgen («acre» = 40 Aren oder 4047 m²). Eine Einheit entspricht also etwa einem Ertrag von ¼ Kilogramm pro Quadratmeter.

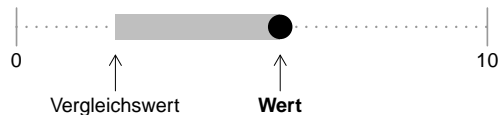
⁵ In Fachwerkgrafiken («trellis displays») werden viele Grafiken gleichzeitig auf einer einzigen Seite dargestellt. – Vgl. Becker et al. [Trellis, 1996]; Becker/Cleveland [Trellis/Man, 1996]; Cleveland [Visualizing, 1993] und <http://cm.bell-labs.com/cm/ms/departments/sia/project/trellis/>. – Für Fachwerkgrafiken gibt es im R-Projekt das Paket «lattice».

D.5 Beschreibung: Punktbalken

Um absolute Werte und deren Differenz darzustellen, können anstelle von Dreiecken auch «Punktbalken» verwendet werden.

Punktbalken⁶ sind auf einer Achse frei positionierbare Balken, deren Länge einer Differenz entspricht. Eines der Balkenenden ist mit einem Punkt markiert. Damit werden ein Wert und ein zugehöriger Vergleichswert unterscheidbar. Zur Darstellung von negativen und positiven Abweichungen kann der Balken unterschiedlich eingefärbt werden.

► Tafel 13



Tafel 13:
Legende zum
Symbol für Punkt-
balkendiagramme

D.6 Beispiel 3: Gerstenernten 1931 und 1932

Auch bei der Abbildung der Daten zu den Gerstenernten mit Punktbalken sieht man die Unterschiede zwischen den Ernten der beiden Jahre gut.

Durch die wegfallende Pfeilform des Dreiecks, wäre die Ausrichtung der Differenz ein wenig schlechter erkennbar, wenn sie nicht durch die unterschiedliche Einfärbung deutlich angezeigt würde.

Dafür ist der Endwert mit dem Punkt sehr deutlich markiert. Dies ermöglicht es, diese Werte über die ganze Erhebung besser zu überblicken.

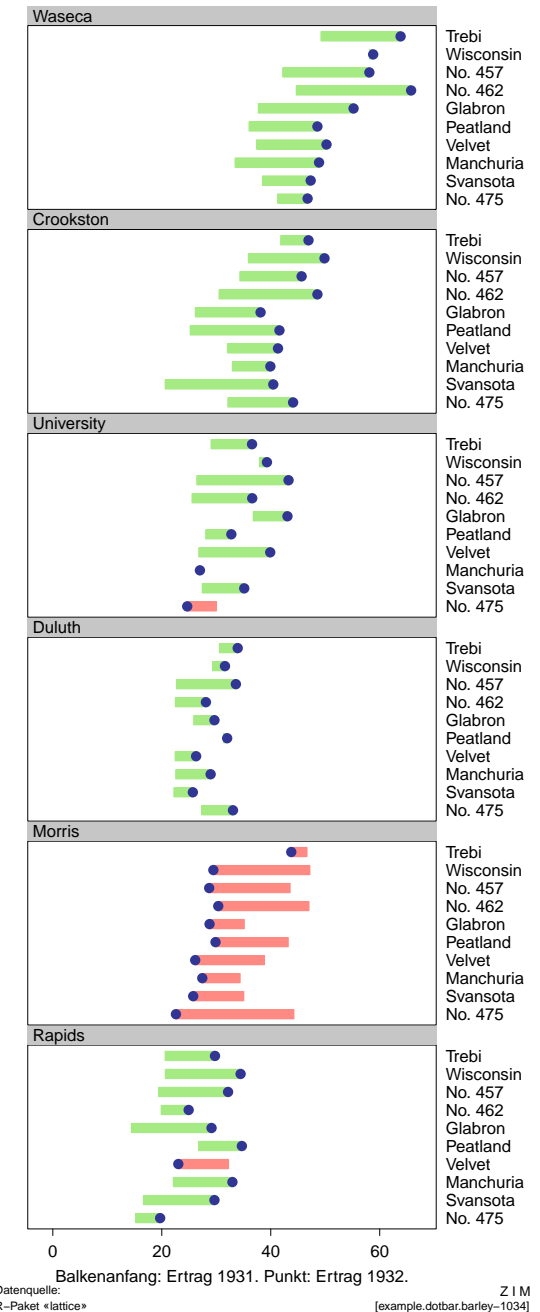
Zur Diskussion

► Tafel 14

↑ Daten: S. 50

⁶ Englisch: «dot bar». – Fischer [Grafiken zur PCS-Beurteilung, 2008]: 44.

Tafel 14:
Gerstenernten in
Minnesota 1931
und 1932 (Punkt-
balkendiagramm)



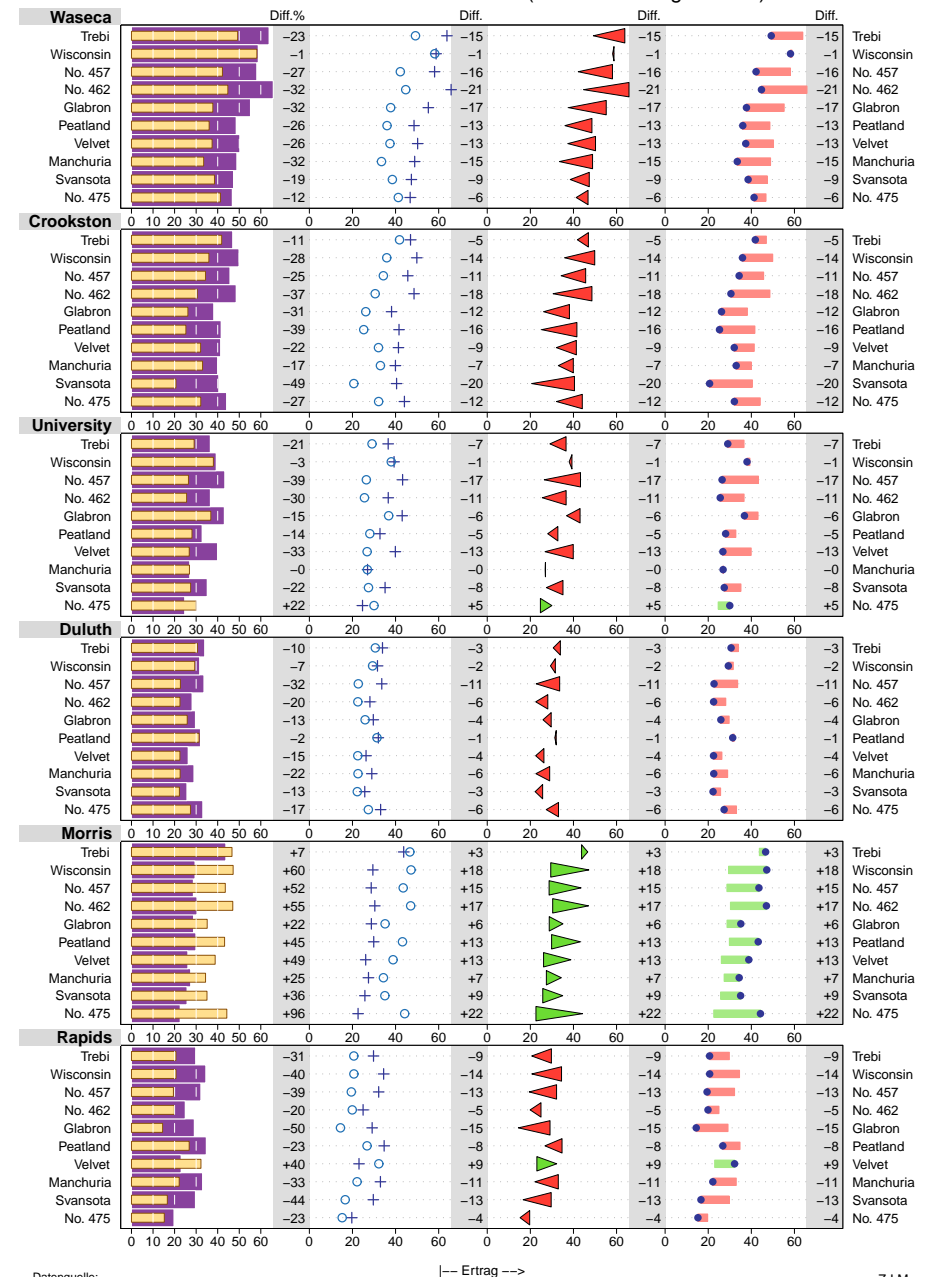
D.7 Vergleich von vier Darstellungsweisen

In der folgenden Grafik werden die Gerstenernte-Daten spaltenweise nach vier unterschiedlichen Darstellungsweisen präsentiert:

► Tafel 15

1. **Balkendiagramm:** Durch die Überlagerung der Balken der Ernten aus den beiden Beobachtungsjahren wird gut sichtbar, ob es eine Abnahme oder eine Zunahme gab.
2. **Punktendiagramm:** Hier kann man den Verlauf der einzelnen Ernten über alle Sorten und Anbauorte besser verfolgen als in den anderen Darstellungsweisen. Dagegen braucht es mehr Beobachtungsaufwand, um die Unterschiede klar zu erkennen.
3. **Differenzdiagramm mit Dreiecken:** Die Unterschiede werden hier besonders augenfällig. Auch der Verlauf zu immer höheren Ernten in den im oberen Bereich abgebildeten Anbaubereichen ist sichtbar.
4. **Differenzdiagramm als Punktbalkendiagramm:** Die Unterschiede sind gut sichtbar. Die Ausprägung der Ernte vom 1932 (beim Punkt) über alle Anbaubereiche und Sorten ist wie im Punktendiagramm gut erkennbar.

Tafel 15: Gerstenernten in Minnesota 1931 und 1932 (vier Darstellungsweisen)



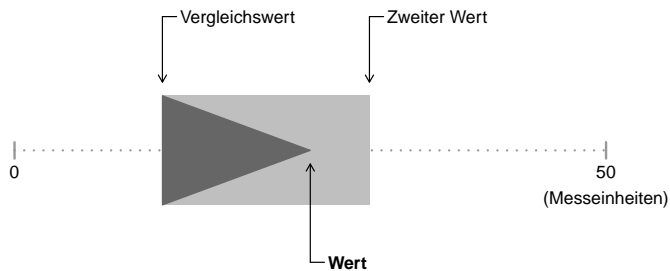
D.8 Beschreibung: Dreieck mit Balken

Zur Anzeige einer weiteren Differenz kann an die Basis des Dreiecks ein Balken angefügt werden. Er kann in die gleiche oder in die dem Dreieck entgegengesetzte Richtung blicken. Auch wenn er hinter dem Dreieck liegt, bleibt er sichtbar, denn das Dreieck verjüngt sich bis zur Spitze, der Balken aber behält immer die gleiche Breite. Wenn der Balken allerdings sehr klein im Verhältnis zum Dreieck ist, verschwindet er trotzdem mehr oder weniger. Abhilfe geschaffen werden kann, indem die vordere Balkenkante über dem Dreieck gezeichnet wird.

► Tafel 16

↑ Beispiel:
Tafel 17 (S. 57)

Tafel 16:
Legende zum
Symbol für
Doppeldifferenz-
diagramme



D.9 Beispiel 4: Gerstenernten 1931 und 1932

Um die Gerstenernten etwas besser beurteilen zu können, wurde auch noch die durchschnittliche Ernte pro Sorte berechnet und angezeigt.

Dazu wurde ein Doppeldifferenzdiagramm verwendet: Das Dreieck zeigt wie bereits früher den Unterschied von der Ernte 1931 zur Ernte 1932. Der dahinterliegende, heller gefärbte Balken zeigt den Unterschied von der Ernte 1931 zur durchschnittlichen Ernte der betreffenden Sorte in diesem Jahr.

Man sieht, dass in 'Waseca' die Ernte von 1931 bei allen Sorten überdurchschnittlich war (lauter hellgrüne Balken). Die Ernte von 1932 war deutlich niedriger als die Ernte von 1931 (rote Dreiecke), aber immer noch etwas höher als die durchschnittliche Ernte pro Sorte von 1931. (Fast alle hellgrünen Balken sind länger als die roten Dreiecke.)

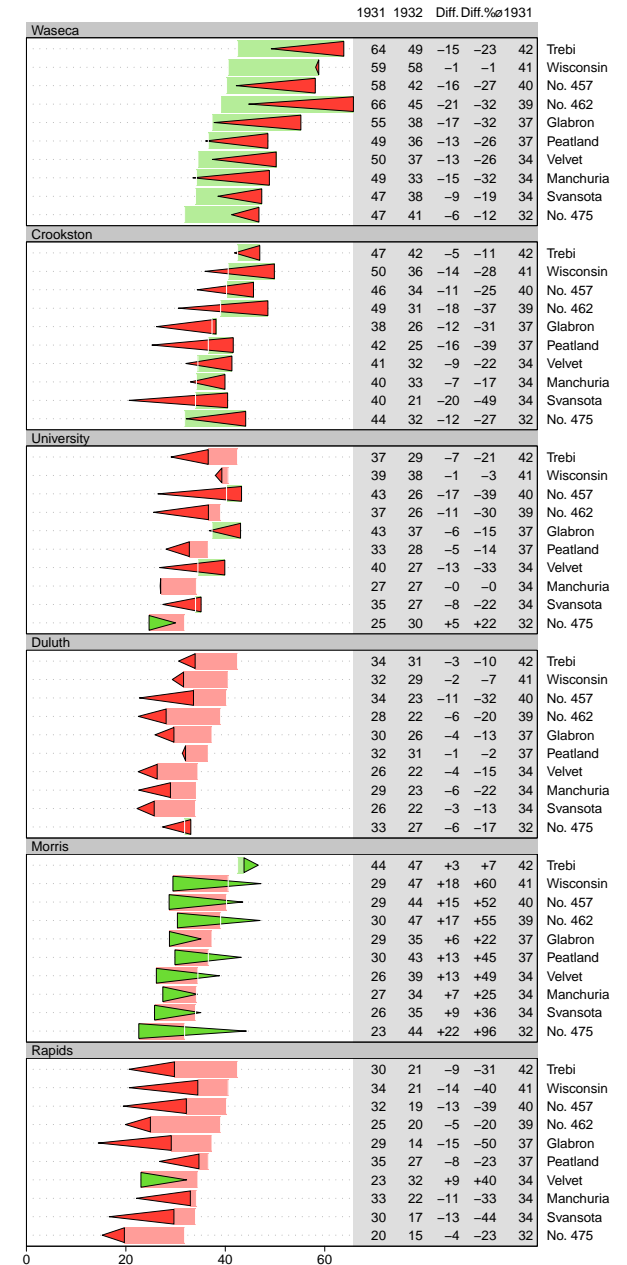
In 'Duluth' und in 'Rapids' waren die Ernten bereits 1931 sehr unterdurchschnittlich. (Bis auf einen sind alle Balken hellrot.)

► Tafel 17

Grafik

Resultate

Tafel 17:
Gerstenernten in
Minnesota 1931
und 1932 (Doppel-
differenzdiagramm)



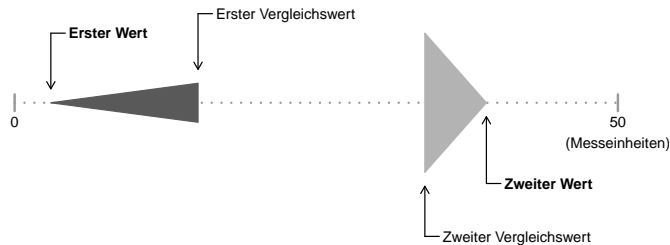
Datenquelle: Dreieck: Basis: Ertrag 1931. Spitze: Ertrag 1932. Balken: ø 1931 pro Sorte.
R-Paket «lattice» [example.diffplot.barley-1034]

Z I M

D.10 Beschreibung: Kombination von zwei Dreiecken

Wenn zwei Differenzen zu unterschiedlichen Ausgangswerten angezeigt werden müssen, können zwei Dreiecke auf der gleichen Zeile miteinander kombiniert werden.

Wenn es nicht zu viele Überlagerungen gibt, kann dies auch mit mehr als zwei Dreiecken gemacht werden.⁷ Zur Unterscheidung der beiden Differenzen kann die Farbintensität verändert werden. Positive und negative Differenzen werden mit kontrastierenden Farben versehen.



Zwei Dreiecke auf einer Zeile
► Tafel 18

Tafel 18:
Legende zum
Symbol für Differenzdiagramme
mit zwei Dreiecken

D.11 Beispiel 5: Gerstenernten 1931 und 1932

Im vorangegangenen Beispiel wurden die Ernten von 1932 mit jenen von 1931 und diese mit den durchschnittlichen Ernten von 1931 pro Gerstensorte verglichen. Noch interessanter als diese Vergleiche ist die Gegenüberstellung der Vergleiche der Ernten von 1931 und 1932 mit den jeweiligen Durchschnittserträgen pro Sorte.

Zur Darstellung dieser vier Werte wurden zwei Dreiecke verwendet. Diese zeigen die Differenzen zwischen Ertrag und durchschnittlichem Ertrag in den beiden Jahren an. Aus der unterschiedlichen Positionierung der beiden Dreiecke wird zusätzlich die Differenz zwischen den Erträgen von 1931 und 1932 ersichtlich.

Es zeigt sich, dass in den vielen Anbauorten die Abweichungen vom durchschnittlichen Ertrag zwar unterschiedlich hoch waren, aber meist in die gleiche Richtung gingen. In 'Waseca' beispielsweise waren die Erträge bei allen Sorten in beiden Jahren überdurchschnittlich, in 'Rapids' und 'Duluth' hingegen waren die Erträge in beiden Jahren bei fast allen Sorten unterdurchschnittlich. Nicht der Fall war dies in 'Morris' und 'Crookston' sowie bei einzelnen Sorten in der 'University Farm'.

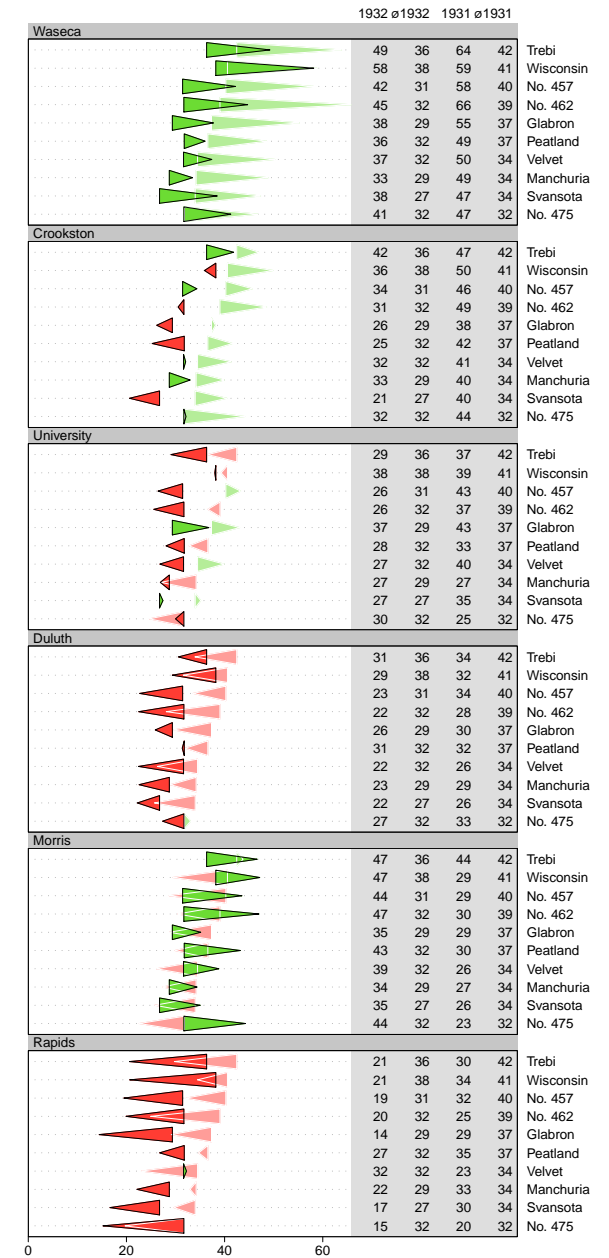
Einleitung

Grafik
► Tafel 19

Resultat

⁷ Ein Beispiel für drei Kennzahlen, die sich gut mit drei nebeneinander gestellten Dreiecken abbilden lassen, sind die durchgängig in aufsteigender Reihenfolge definierten DRG-Kennzahlen 'untere Grenzverweildauer', 'mittlere Verweildauer', 'obere Grenzverweildauer'. – Vgl. Fischer [Grafiken zur PCS-Beurteilung, 2008]: 40 + 90 f.

Tafel 19:
Gerstenernten in
Minnesota 1931
und 1932 (zwei Differenzdiagramme)



Dreieck im Vordergrund: Spitze: Ertrag 1932, Basis: \bar{x} 1932 pro Sorte.

Datenquelle: Dreieck im Hintergrund: Spitze: Ertrag 1931, Basis: \bar{x} 1931 pro Sorte.

R-Paket «lattice»

Z I M
[example.diffplot.barley-1034]

Wenn die Werte von «Morris» – wie von Cleveland vorgeschlagen⁸ – durch Umtauschen korrigiert werden, verändern sich nicht nur die Werte von «Morris» selbst, sondern auch die Durchschnittswerte pro Gerstensor- te und damit alle Differenzen.

Es gibt immer noch drei Anbauorte mit gleichartigen Differenzen: «Waseca» mit lauter positiven, «Duluth» und «Rapids» mit fast lauter negativen Differenzen.

Mit den korrigierten Werten ist «Morris» ins mittlere Feld mit uneinheitlicher Ausrichtung gerutscht. Auffällig sind hier nun die Sorten «Wisconsin», «No. 475» und «Velvet», die von überdurchschnittlichen Erträgen im Jahr 1931 auf deutlich unterdurchschnittliche Erträge im Jahr 1932 fielen.

Ins Auge springen auch die unerwarteten überdurchschnittlichen 1932er-Erträge von «Peatland» in «Duluth» und von «Velvet» in «Rapids». Die Auffälligkeit von «Velvet» in «Rapids» wurde in der Literatur bereits diskutiert;⁹ von «Peatland» in «Duluth» berichtet Cleveland aber nichts.

Wenn man es sich genauer überlegt, ist der Fall von «Velvet» in «Rapids» aus der Sicht dieser Grafik deshalb von Interesse, weil sich beim Vertauschen der Jahre – d. h. der Spitzen der Dreiecke – zwei gleich ausgerichtete, rote Dreiecke ergäben, die mit den restlichen Dreiecken harmonischer korrespondieren würden. Das wäre bei «Peatland» in «Duluth» nicht so; die beiden Dreiecke würden sogar noch grösser. Die Auffälligkeit besteht in diesem Fall darin, dass hier die einzige positive (grüne) Differenz im ganzen Grafikfeld auftritt.

Als Datendetektiv kann man weitere Fälle nach dem Muster von «Velvet» in «Rapids» suchen: Wo treten Überlagerungen von Dreiecken mit unterschiedlicher Ausrichtung auf? Auf diese Weise entdeckt man noch «No. 475» auf der «University Farm».

Die Darstellung von zwei nebeneinander gestellten Dreiecken ermöglicht die Suche nach Beziehungsmustern zwischen den beiden Dreiecken und löst so Fragen aus, die weitere Überlegungen und Entdeckungen mit sich bringen. Damit wird ein wichtiges Ziel der Datenvisualisierung erfüllt.

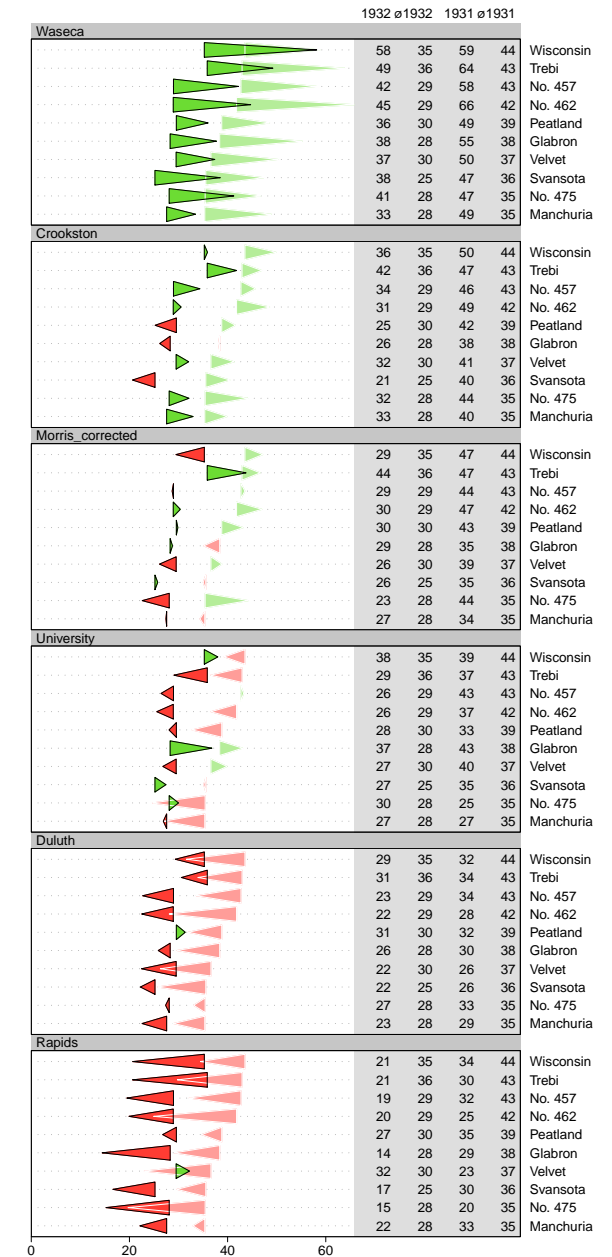
Korrektur der Werte von «Morris»

► Tafel 20

Auffälligkeit von «Velvet» in «Rapids»

Konklusion

Tafel 20:
Gerstenernten in
Minnesota 1931
und 1932, korrigiert
(zwei Differenzdia-
gramme)



Dreieck im Vordergrund: Spitze: Ertrag 1932, Basis: \bar{x} 1932 pro Sorte.
Datenquelle: Dreieck im Hintergrund: Spitze: Ertrag 1931, Basis: \bar{x} 1931 pro Sorte.
R-Paket «lattice» [example.diffplot.barley-1034] Z I M

⁸ Cleveland [Visualizing, 1993]: 329 ff.

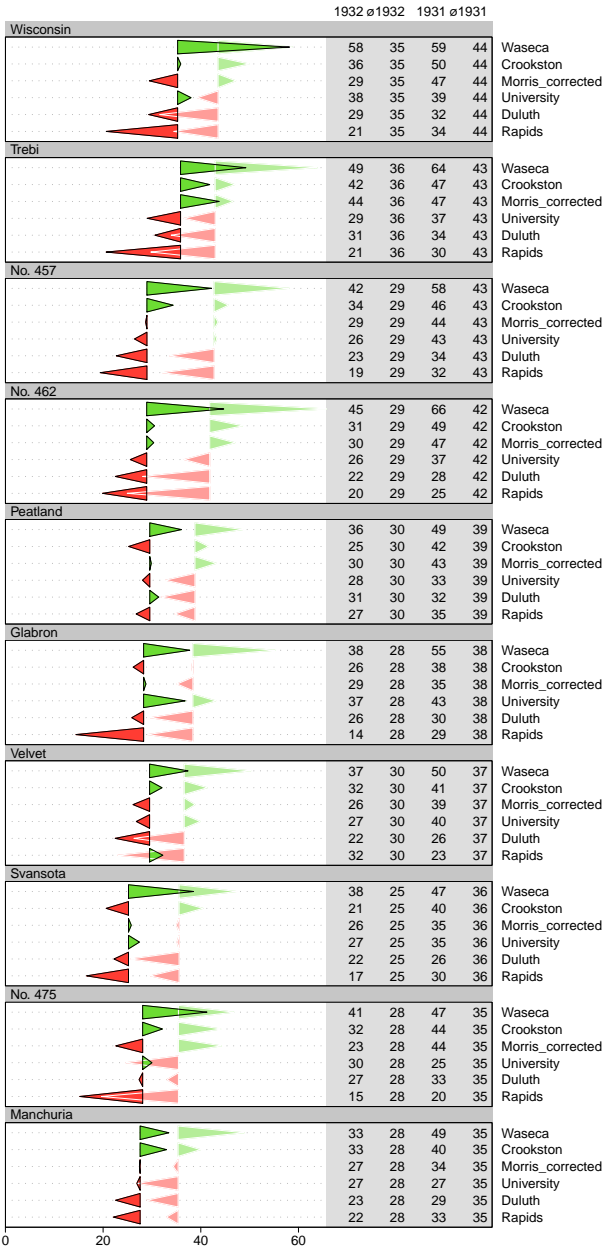
⁹ Vgl. Cleveland [Visualizing, 1993]: 332+340.

Durch Umstellen der Grafik von Grafikfeldern pro Anbauort auf Grafikfelder pro Gerstensorte können die Abweichungen von den durchschnittlichen Erträgen sehr augenfällig dargestellt werden.

Die grosse Bandbreite der Differenzen bei allen Sorten führt zur Vermutung, dass die Mehr- und Mindererträge stärker von den Anbauorten als von der Sorte abhängig sind. Dazu wurde eine weitere Grafik erstellt.

Umstellung zu Auswertungen pro Gerstensorte
► [Tafel 21](#)

Tafel 21:
Gerstenernten in Minnesota 1931 und 1932, korrigiert (zwei Differenzdiagramme), nach Sorten



Dreieck im Vordergrund: Spitze: Ertrag 1932, Basis: ø 1932 pro Sorte.
Datenquelle: Dreieck im Hintergrund: Spitze: Ertrag 1931, Basis: ø 1931 pro Sorte.
R-Paket «lattice» [example.diffplot.barley-1034] Z 1 M

In dieser Grafik wurden anstelle der Abweichungen von den Durchschnitten pro Gerstensorte (und Jahr) die Abweichungen von den Durchschnitten pro Anbauort (und Jahr) dargestellt.

Man sieht hier – im Unterschied zur vorangegangenen Darstellung –, dass die Dreiecke im Durchschnitt kürzer und damit die Abweichungen vom durchschnittlichen Ertrag des Anbauortes und Jahres kleiner sind. Dies zeigt, dass – wie oben bereits vermutet – die Mehr- und Mindererträge stärker von den Anbauorten als von der Sorte abhängig sind.

Diese Sicht kann rechnerisch belegt werden: Die Reduktion der absoluten Abweichungen vom Median ($r1$) beträgt bei der Gruppierung nach Anbauorten und Jahren 52 %, bei der Gruppierung nach Gerstensorten und Jahren hingegen nur 4 %.¹⁰

Die weiter oben vermerkte Auffälligkeit von 'Velvet' in 'Rapids' mit den überlagerten gegenläufigen Dreiecken ist auch hier sichtbar, ebenso der Fall von 'No. 475' auf der 'University Farm'.

Grafik

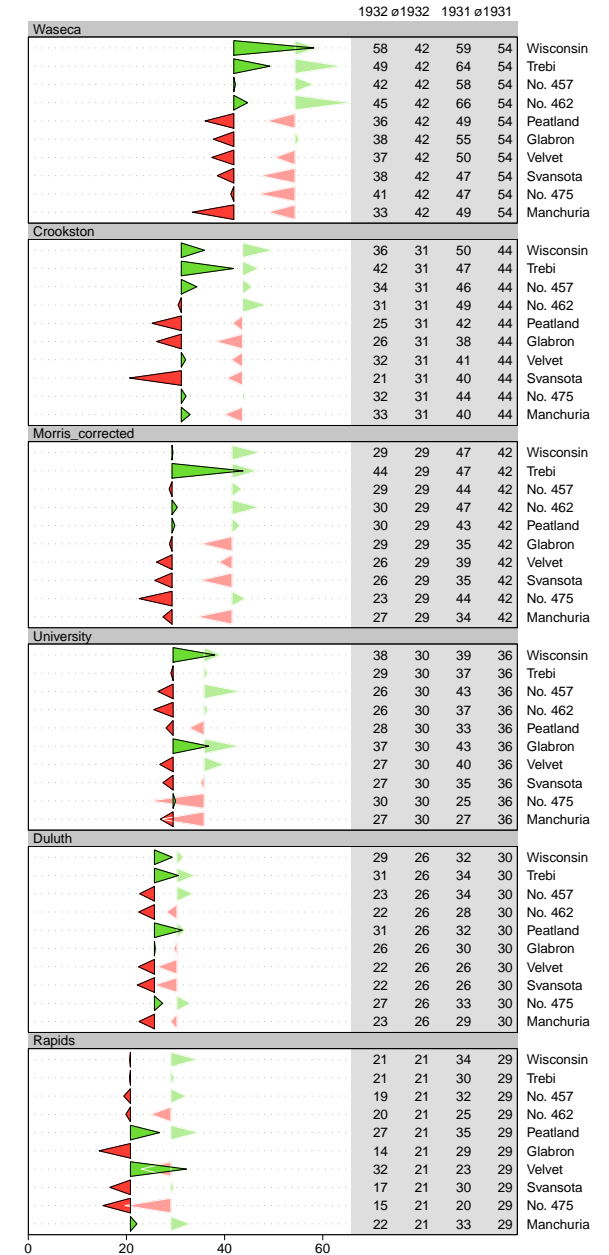
► Tafel 22

Resultate

↑ $r1$: S. 102

↑ S. 60

Tafel 22:
Gerstenernten in
Minnesota 1931
und 1932, korrigiert
(zwei Differenzdia-
gramme), nach
Anbauorten



¹⁰ Die Varianzreduktion ($r2$) beträgt bei der Gruppierung nach Anbauorten und Jahren 73 %, bei der Gruppierung nach Gerstensorten und Jahren hingegen nur 1 %.

Um noch auf eine andere Weise zu sehen, ob die Mehr- und Mindererträge stärker vom Anbauort oder von der Sorte abhängig sind, wurde zuguterletzt noch ein Strömungsbild erstellt.

Als Erstes fallen die grundsätzlich dunkleren Farben im Grafikfeld für 1931 auf. Sie zeigen die insgesamt höheren Ernten im Jahr 1931 an. Im Weiteren ist eine Ertragssteigerung von «Rapids» bis «Waseca» zu beobachten (mit Farben, die von gelb/grün zu blau/violett wechseln).

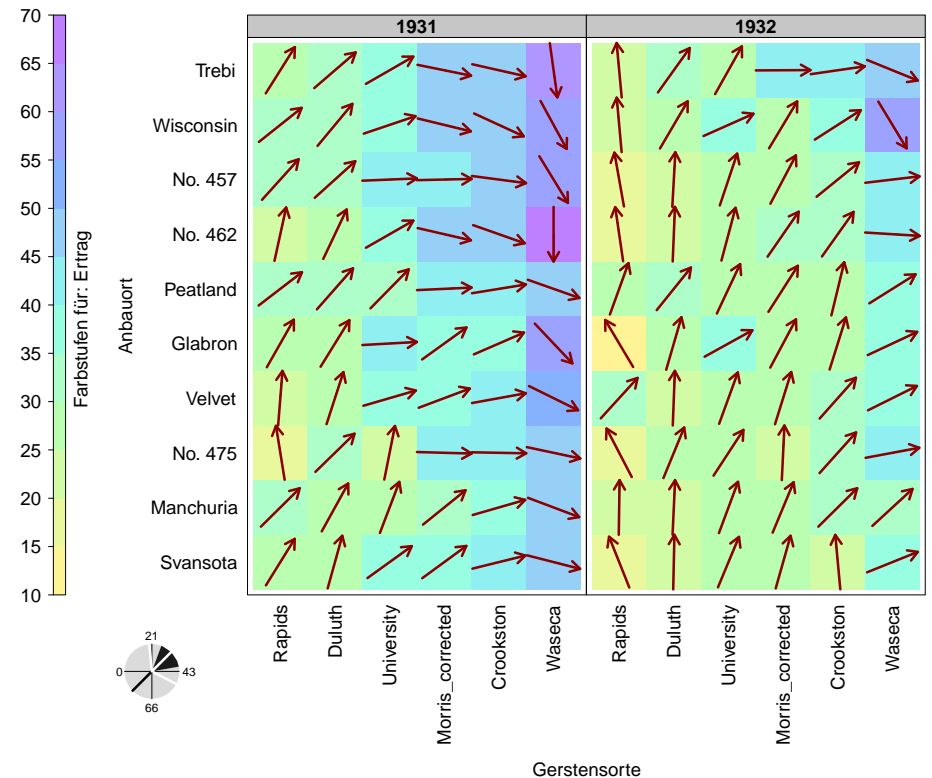
Bei näherem Hinschauen zeigt sich, dass die Schwankungen bei vertikalem Vergleich der Ausrichtungen der Pfeile (und der Farben) kleiner sind als die Schwankungen in horizontaler Richtung: Während es in den einzelnen Spalten zwei bis drei Farbausprägungen gibt, sind auf den Zeilen Spannweiten von meist drei bis vier Farbeintragungen zu finden. Dies bestätigt die Vermutung, dass die Mehr- und Mindererträge weniger von der Sorte als vielmehr vom Anbauort abhängig sind.

Strömungsbild

► Tafel 23

Resultate

Tafel 23: Gerstenernten in Minnesota 1931 und 1932, korrigiert (Strömungsbilder)



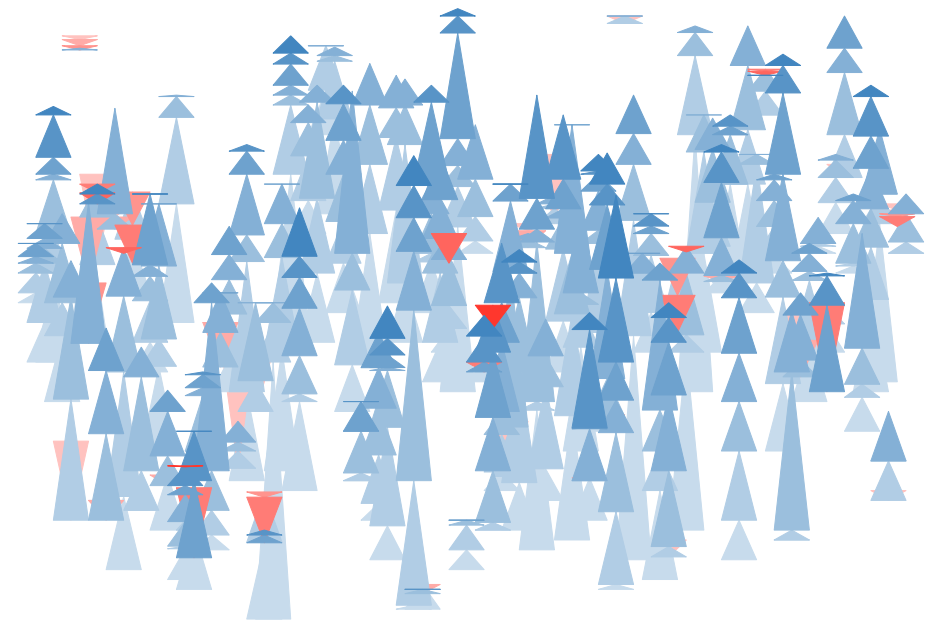
Streuungsflächen links: Der schwarze Fächer zeigt die Werte der Variablen "Ertrag" vom ersten bis zum dritten Quartil auf einer Kreisskala von 0 bis 66.

Die weisse Federn zeigen die Positionen des 5 %- und des 95 %-Perzentils.

Datenquelle:
R-Paket «lattice»

Z I M
[example.vanechart2.barley~1034]

E Sequenzdiagramme



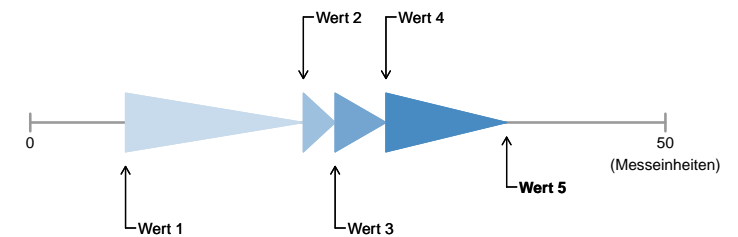
Rot verloren im Blauwald (2010)

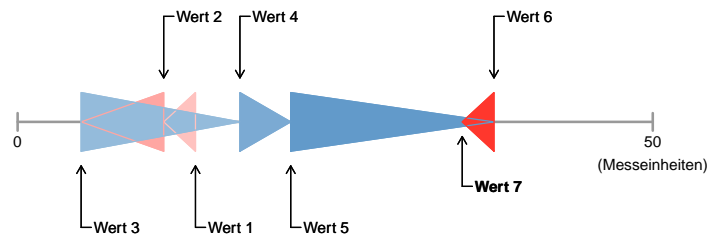
Rotierter Ausschnitt aus einem Sequenzdiagramm zu Rehabilitations-
behandlungen, mit Vergrößerung der Symbolbreite. ↑ Vgl. Tafel 27 (S. 77)

E.1 Motivation

Liniendiagramme?	Zur Darstellung der Entwicklung einer Kennzahl im Zeitverlauf bieten sich Liniendiagramme an: Auf der x-Achse wird die Zeit abgetragen und auf der y-Achse der Wert. Liegen Werte zu mehreren Untersuchungseinheiten vor, können mehrere Liniendiagramme erstellt und untereinander oder nebeneinander angeordnet werden. In beiden Fällen ist ein Vergleich nur behelfsmässig möglich. Ausserdem brauchen die zweidimensionalen Liniendiagramme relativ viel Platz.
Viele Untersuchungseinheiten	Es stellte sich mir die Frage, wie ich Verläufe von Kennzahlen zu vielen Untersuchungseinheiten so darstellen könnte, dass sie gut vergleichbar sind und auf möglichst einer Druckseite Platz haben. Es waren z. B. Veränderungen von Fähigkeitseinschränkungen vieler Patienten über mehrere Wochen oder Veränderungen von Kosten und Erträgen zu vielen Patientenkategorien über mehrere Jahre hinweg übersichtlich darzustellen.
Sequenzierte Dreiecke	Wie im vorherigen Kapitel beschrieben, lassen sich mit Dreiecken Differenzen darstellen. Da Verläufe über mehrere aufeinanderfolgende Zeitperioden Sequenzen von Differenzen sind, überlegte ich mir, ob ich solche Verläufe nicht mit Sequenzen von Dreiecken abbilden könnte? Der Vorteil wäre, dass dies eine eindimensionale Darstellung ermöglichte. Der Nachteil wurde mir allerdings schnell auch klar: Was passiert, wenn sich das Vorzeichen der Veränderung wechselt? Die aufeinander folgenden Dreiecke überlagern sich.
Umgang mit Richtungswechseln	Trotz dieses Problems erstellte ich einige Versuchsgrafiken. Ich entdeckte, dass das Problem der Überlagerung gemindert wurde, wenn ich Füllung und Rahmen der Dreiecke in zwei Durchgängen aufzeichnete: Damit wurde der Rahmen eines überlagerten Dreiecks über der Füllung des darüber liegenden Dreiecks sichtbar. Pendelbewegungen konnten damit besser erkannt werden. Überlagerungen brachten es zusätzlich mit sich, dass Anfang und Ende der Sequenzen z. T. nicht mehr gut erkenntlich waren. Eine Milderung dieses Problems fand ich, indem ich die Farbintensität vom Beginn einer Sequenz bis zu deren Ende zunehmen liess.

Tafel 24:
Legende zu
sequenzierten
Dreiecken
mit fünf Werten





Tafel 25:
Legende zu
sequenzierten
Dreiecken
mit sieben Werten

Aneinandergereihte
Dreiecke

► Tafeln 24 bis 26

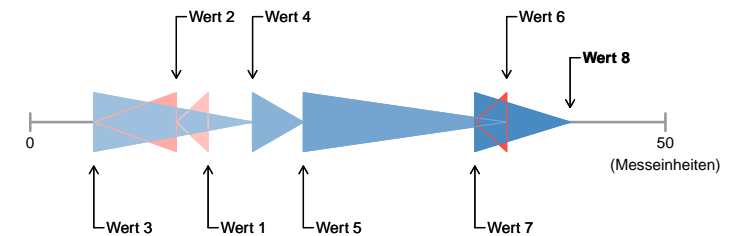
Farben und
Richtungswechsel

E.2 Beschreibung: Sequenzierte Dreiecke

In einem Sequenzdiagramm¹ werden aufeinander folgende Veränderungen von Variablenwerten durch grafische Symbole angezeigt, die auf einer Linie aufgereiht sind. Als Symbole bieten sich Dreiecke an: Die Spitze eines Dreiecks wird zur Basis des nächsten Dreiecks und damit zum Ausgangspunkt für die nächste Differenz. Dank dieser kompakten Form wird es möglich, Positionen und Veränderungen zu mehreren Untersuchungseinheiten gleichzeitig darzustellen.

Da sich die dargestellten Werte in freiem Wechsel sowohl erhöhen als auch senken können, wurde einerseits die Veränderungsrichtung mit zwei deutlich unterschiedlichen Farben gekennzeichnet und andererseits die Farbintensität von Messpunkt zu Messpunkt verstärkt.

Tafel 26:
Legende zu
sequenzierten
Dreiecken
mit acht Werten



¹ Fischer [Krankenhaus-Betriebsvergleiche, 2005]: 108f.

E.3 Beispiel 1: Verlauf von Rehabilitationsbehandlungen

In der Grafik wird eine Stichprobe von Rehabilitationspatienten gezeigt, welche zwischen fünf und sieben Wochen in stationärer Behandlung waren. Die Patienten sind pro Zeile eingetragen. Die (anonymisierte) Patientenidentifikation steht zu Beginn der Zeile. Im Grafikfeld kann die Entwicklung der Fähigkeiten gemäss dem FIM-Instrument² abgelesen werden: Für jede Hospitalisationswoche wurden die durchschnittlichen FIM-Punkte eingezeichnet und durch Dreiecke miteinander verbunden. Die Länge der Dreiecke zeigt somit die Verbesserung oder Verschlechterung in FIM-Punkten pro Woche. Die FIM-Punkte jeder Woche wurden zusätzlich als Zahlenwerte in der grauen Spalte rechts eingetragen.

Die Eintragungen der Patienten sind geordnet nach den FIM-Punkten. Zuerst sind die selbständigsten Patienten, zuunterst die unselbständigsten Patienten.

Auf den ersten Blick fällt auf, dass die meisten Dreiecke blau sind, d. h., dass die Fähigkeiten gemäss FIM bei den allermeisten Behandlungen aus der Stichprobe im Verlauf der Behandlung zunahmen. Die Zunahmen sind unterschiedlich hoch. Die grössten Veränderungen sind in der Mitte zu finden; offenbar besteht ein besonders grosses Veränderungspotential bei Patienten, bei welchen zu Beginn der Behandlung ein FIM-Wert von 3 bis 4 Punkten gemessen wurde.

Es gibt nur eine Behandlung bei durchwegs abnehmenden FIM-Werten: die zuunterst eingetragene.

Bei den Behandlungen mit zeitweilig abnehmenden FIM-Werten fällt auf, dass die Abnahme des Öftern am Schluss der Behandlung auftrat. Dies zeigen die intensiv rot gefärbten Dreiecke am Ende einiger Sequenzen.

Zeitweilige Stagnationen der eingetragenen Werte sind aus technischen Gründen nicht erkennbar. Bei Patient «1560» (auf der siebentuntersten Zeile) beispielsweise blieb der FIM-Wert während der ersten drei Behandlungswochen bei 2.72 stehen. Diese repetitierten Werte erscheinen in der Tabelle; in der Grafik sind sie nicht sichtbar.

Es fällt auf, dass die Veränderungen bei mittleren FIM-Werten grösser sind als am Rand der Skala. Dies könnte auch darauf hinweisen, dass die FIM-Skala nicht linear ist, sondern – wie mit einer Rasch-Analyse belegt wurde³ – für solche Auswertungen linearisiert werden müsste.

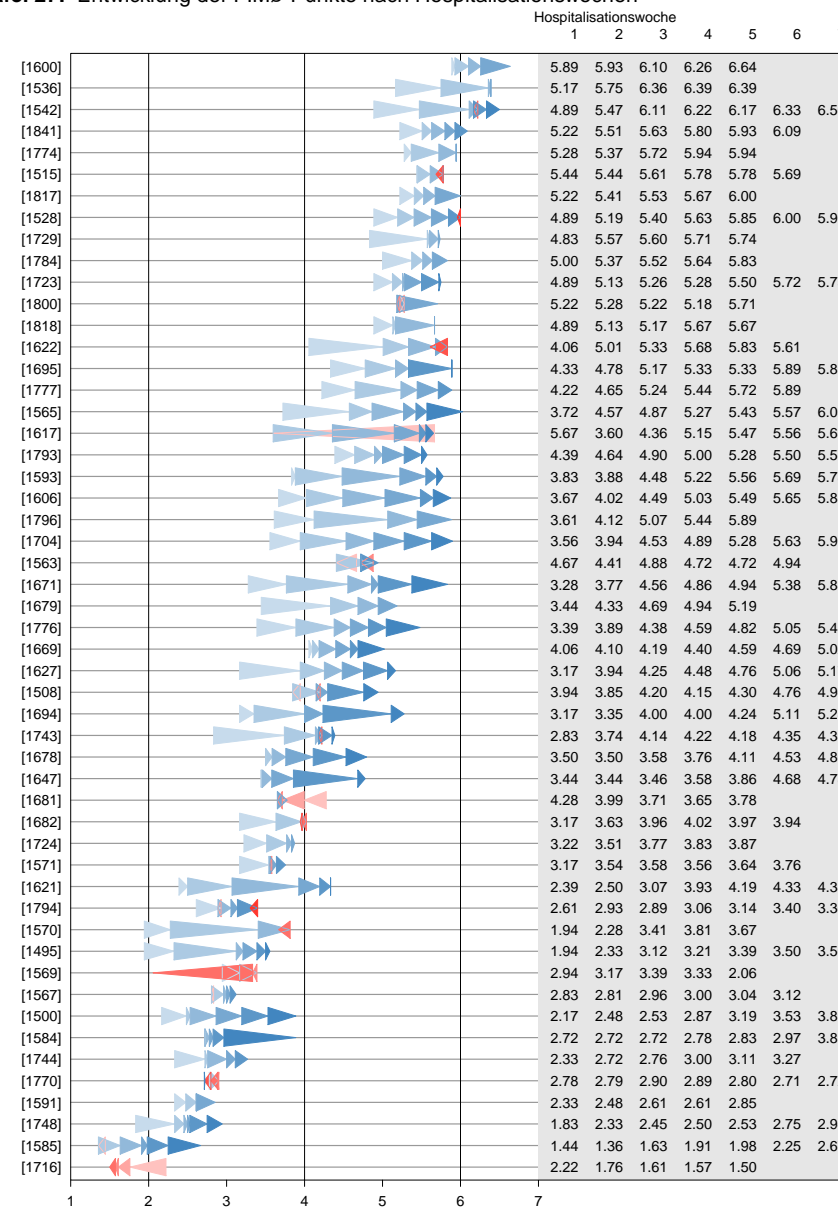
Grafik

► Tafel 27

Resultate

Zur Diskussion

Tafel 27: Entwicklung der FIM-Punkte nach Hospitalisationswochen



--- FIM-Wert ---
 Hellere Pfeile: ältere Veränderungen. Dunklere Pfeile: neueste Veränderungen.
 Färbung: bläulich -> Verbesserung; rötlich -> Verschlechterung.

Datenquelle:
 Test-Daten SZH

52 Behandlungsfälle

Z I M
 [SZH.08Z.seqplot.tri-1033]

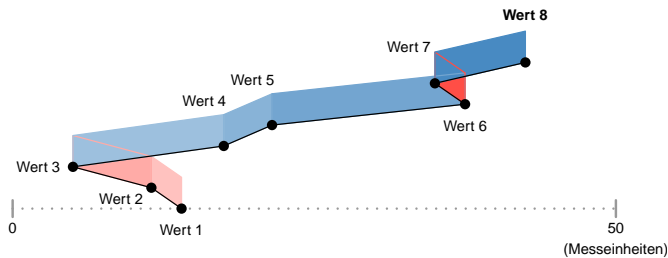
² FIM = Functional Independence Measure. Die FIM-Punkte wurden berechnet als FIM-Punktesumme dividiert durch die Anzahl beurteilter FIM-Items. Die Skala der FIM-Punkte reicht von «1», wenn eine totale Hilfestellung nötig ist, bis «7» bei völliger Selbstständigkeit. – Granger et al. [UDSmr-93, 1995]. <http://www.udsmr.org/>.

³ Vgl. Granger et al. [FIM-Rasch, 1993].

E.4 Beschreibung: Sequenzbänder

Anstelle von Dreiecken können auch Bänder über einer Werteskala ausbreitet werden. Damit die Differenzen von einem Messpunkt zum nächsten besser erkannt werden können, wurde jeder Messpunkt mit einem Punkt grafisch markiert und die Differenzen selbst immer intensiver gefärbt. Zusätzlich wurde das in waagrechter Richtung verlegte Band von einem Messpunkt zum nächsten um einen bestimmten, kleinen Betrag senkrecht versetzt. Somit erscheinen positive und negative Differenzen wie ineinander gefaltet. Zur besseren Lesbarkeit wurden für positive und negative Differenzen deutlich unterscheidbare Farben verwendet.

Sequenzbänder benötigen etwas mehr Platz als sequenzierte Dreiecke, zeigen aber Verlaufswechsel besser.



► Tafel 28

Tafel 28:
Legende zu einem
Sequenzband
mit acht Werten

E.5 Beispiel 2: Verlauf von Rehabilitationsbehandlungen

In der Grafik sind die gleichen Daten aus der Rehabilitation verwendet worden wie in der vorangegangenen Grafik mit den sequenzierten Dreiecken. Der Lesbarkeit wegen werden aber nur zwei Drittel der Behandlungsfälle gezeigt.

Es gelten die gleichen Aussagen wie zur vorangegangenen Auswertung. Die Fälle mit zeitweilig abnehmenden FIM-Werten sind nun aber besser sichtbar. Zum Beispiel sah man beim Patient «1561» (auf der zweituntersten Zeile) vorher nicht so deutlich, dass sich der Zustand von der ersten zur zweiten Woche verschlechterte. Auch der wechselhafte Verlauf bei Patienten wie «1539» (leicht oberhalb der Mitte) oder «1545» (auf der siebtuntersten Zeile) kann hier besser abgelesen werden. Die Stagnation der Werte während der ersten drei Wochen der Behandlung von Patient «1560» (viertunterste Zeile) fällt erst hier auf.

Der Entscheid für den Einsatz von sequenzierten Dreiecken oder von Sequenzbändern ist also abhängig davon, ob man lieber möglichst viele Fälle oder eine etwas detaillierte Sicht auf die Daten zeigen will.

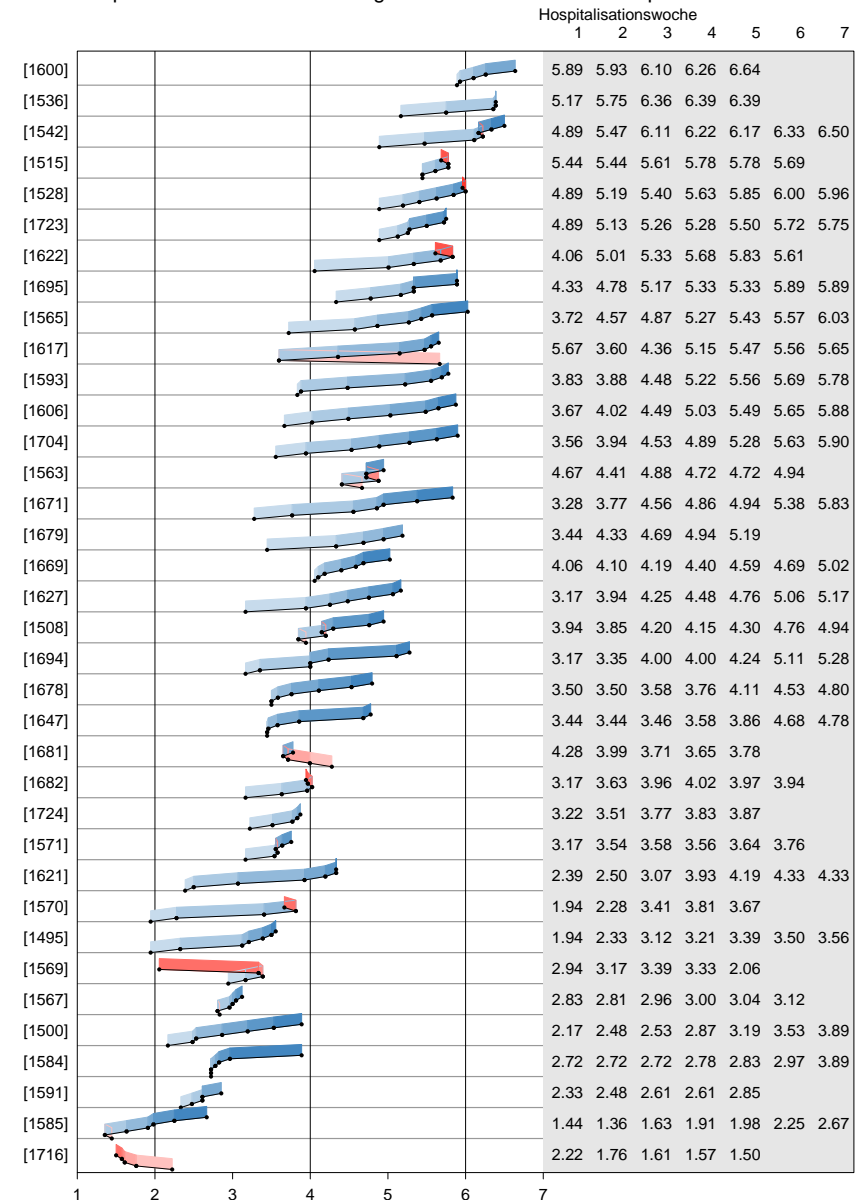
Grafik

► Tafel 29

Resultate

Konklusion

Tafel 29: Sequenzbänder zur Entwicklung der FIM-Werte nach Hospitalisationswochen



--- FIM-Wert --->
Hellere Bandabschnitte: ältere Veränderungen. Dunklere Bandabschnitte: neuste Veränderungen.
Färbung: bläulich -> Verbesserung; rötlich -> Verschlechterung.

Datenquelle:
Test-Daten SZH

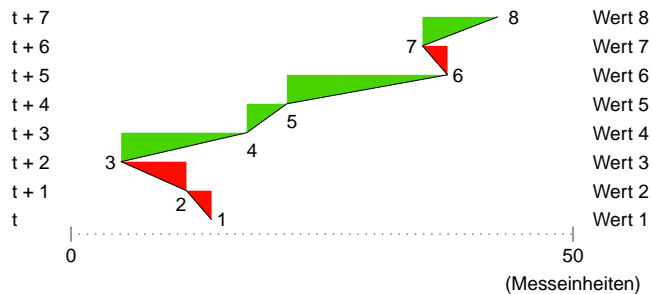
36 Behandlungsfälle

ZIM
[SZH.082.seqplot.para-1033]

E.6 Beschreibung: Kammlinien

Man kann die vertikale Versetzung von Sequenzbändern soweit vergrößern, bis von einem Messpunkt zum nächsten eine ganze beschreibbare Zeile liegt. Wenn man die Bandabschnitte zwischen den Messpunkten nun durch rechtwinklige Dreiecke ersetzt, entsteht eine «Kammlinie». Die Breite der Dreiecke zeigt die Differenz an. Die Höhe ist konstant und wird idealerweise so gewählt, dass sie der Höhe einer Schreibzeile entspricht.

Damit ist man im Prinzip wieder beim Liniendiagramm angelangt. Nur geht die Linie nun nicht von links nach rechts, sondern von unten nach oben. Das hat insbesondere den Vorteil, dass jeder Messpunkt beschriftet werden kann, ohne dass der Text rotiert werden muss. Im Unterschied zum gewöhnlichen Liniendiagramm sind in Kamminiendiagrammen die Differenzen dank der Dreiecke und deren Färbung besser erkennbar.



► Tafel 30

Tafel 30:
Legende zu einer
Kammlinie
mit acht Werten

E.7 Beispiel 3: Verlauf von Rehabilitationsbehandlungen

Die Grafik beruht nochmals auf den gleichen Daten aus der Rehabilitation wie die beiden vorangegangenen Grafiken. Der unselbständigste Patient ist links unten eingetragen. Von unten nach oben und von einer Spalte zur nächsten sind immer selbständigere Patienten zu finden.

Dank der besseren vertikalen Auflösung sind die Behandlungsverläufe hier besonders gut erkennbar. Insbesondere sind in dieser Darstellung auch zeitweilig stagnierende Zustände gut sichtbar. (Vgl. z. B. Patient «1560» in der ersten Spalte auf der siebten Zeile.)

Damit alle 52 Behandlungsfälle auf einer Seite Platz hatten, musste eine für dieses Ausgabeformat zu kleine Schrift gewählt werden.

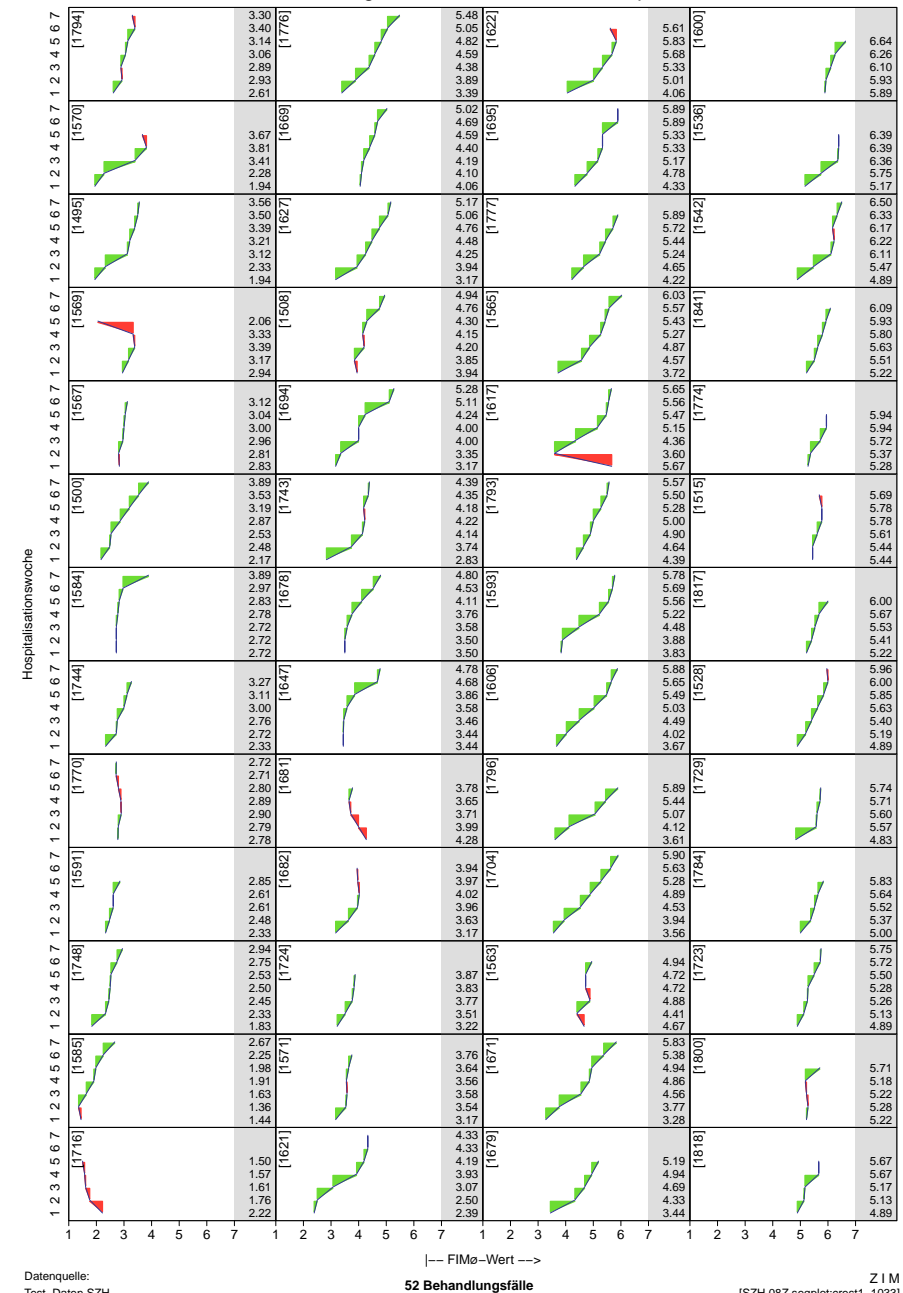
Grafik

► Tafel 31

Resultate

Zur Diskussion

Tafel 31: Kammlinien zur Entwicklung der FIMØ-Punkte nach Hospitalisationswochen



Datenquelle:
Test-Daten SZH

--- FIMØ-Wert --->
52 Behandlungsfälle

Z I M
[SZH.08Z.seqplot:crest1-1033]

F Wechseldiagramme

F.1 Motivation

Paare von Ja/Nein-Antworten	<p>Dreiecke können auch noch andere als wertmässige Differenzen zeigen. Dies entdeckte ich, als ich einen Datensatz mit Vorher-/Nachher-Antworten zu Ja/Nein-Fragen auszuwerten hatte. Aus den Paaren von Antworten können folgende vier Paare von ersten und zweiten Antworten gebildet werden:</p> <ul style="list-style-type: none"> • ja / ja • nein / ja • ja / nein • nein / nein
2×2-Kreuztabelle?	Die einfachste Auswertung einer Erhebung mit solchen Antworten ist eine 2×2-Kreuztabelle, in welche die (absoluten oder relativen) Häufigkeiten pro Antwortpaar eingetragen sind.
Kombinationen von Antwortpaaren	Nun waren im Datensatz aber Antworten zu mehreren Fragen enthalten. Grundsätzlich hätte ich aus den Antworten zu jeder Frage eine eigene Kreuztabelle erstellen können. Da mich aber auch die Kombinationen der Antworten zu allen Fragen pro befragte Person interessierten, suchte ich nach einer Darstellungsmöglichkeit, in der ich jede Antwort einzeln abbilden konnte, die aber so kompakt war, dass ich bei Bedarf mehrere hundert Befragungen auf einer einzigen Seite darstellen konnte. Also galt es, ein grafisches Symbol zu finden, mit dem zwei Zustände und zwei Zustandswechsel abgebildet werden können.
Diagonal unterteilte Quadrate	Nach einigem Herumsuchen blieb ich bei der Flagge hängen, die als Postenmarkierung für Orientierungsläufe benutzt wird: Ein orange-weiss eingefärbtes, diagonal unterteiltes Quadrat. Ich begann, die beiden Dreiecke als dynamische Figuren zu sehen: Das untere Dreieck ist eine Figur, die sich von links nach rechts öffnet, währenddem sich das obere Dreieck von links nach rechts schliesst. Mit dem oberen Dreieck könnte also die erste, «vorherige», nun nicht mehr gültige Antwort codiert werden, mit dem unteren die zweite, «neue» Antwort. Wenn keine Veränderung stattfand, dürfte nach dieser Logik das Quadrat nicht unterteilt werden. Damit dies möglich wurde, gab ich den Antworten «ja» und «nein» unterschiedliche Farben, nämlich Rot und Blau. (Zu dieser Farbenwahl stellte ich mir vor: «ja» = Rot = Feuer und Flamme sein; «nein» = Blau = Es lässt einen kalt.)

Tafel 32:
Legende zu den
Wechselflaggen



F.2 Beschreibung: Wechselflagge

Eine «Wechselflagge» ist ein Quadrat mit einer diagonalen Unterteilung. Damit können zwei Zustände und zwei Zustandwechsel angezeigt werden. Sie eignen sich z. B. zur Abbildung der Kombination von zwei aufeinander folgenden Ja/Nein-Antworten.

Wechselflagge
► Tafel 32

Jeder Antwort wird eine Farbe zugeordnet. (Im Beispiel ist es Rot für «ja» und Blau für «nein».) Die erste Antwort wird hell eingetragen. Wenn die zweite Antwort gleich ausgefallen ist wie die erste, wird sie auch hell eingetragen, und es ergibt sich ein hell gefärbtes Quadrat. Wenn die zweite Antwort anders ist, wird das zweite Dreiecke mit grösserer Farbtintensität eingefärbt.

Farben

Die Antwortpaare können von «ja» bis «nein» geordnet werden, indem man z. B. die zweite Antwort als gewichtiger definiert. Die Reihenfolge lautet dann: «ja / ja», «nein / ja», «ja / nein», «nein / nein».

Reihenfolge

F.3 Beispiel 1: Antwortpaare zu Ja/Nein-Fragen nach Ländern

111 Personen aus fünf Ländern wurden vor und nach der Teilnahme an einem Kurs zu den gleichen sieben Fragen «(a)» bis «(g)» befragt. Die Antwortmöglichkeiten waren «ja» oder «nein». Daraus ergeben sich vier mögliche Paare von erster und zweiter Antwort: «ja / ja», «nein / ja», «ja / nein», «nein / nein».

Einleitung

In der Grafik wurden die Antwortpaare in Form von Wechselflaggen dargestellt. Sie wurden für jedes Land pro Frage gemäss der Reihenfolge in obiger Liste geordnet und nacheinander auf den Zeilen eingetragen.

Grafik
► Tafel 33

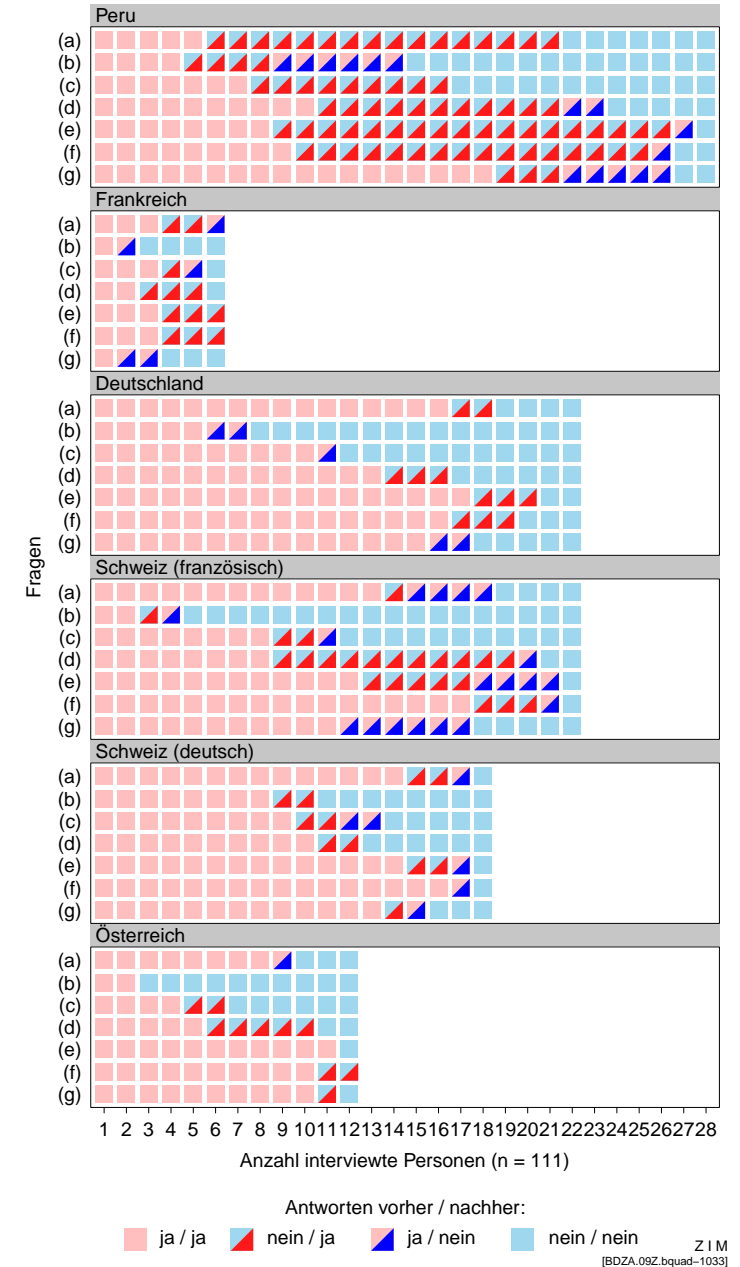
Es fällt zuerst auf, wie sich die Länder nach den Farbanteilen unterscheiden. Dies zeigt, dass das Antwortverhalten in den Ländern unterschiedlich war. Unter den Deutschschweizern und bei den Deutschen gab es am meisten Personen, die schon immer ja sagten. (Es gibt hier anteilmässig mehr hellrote Quadrätchen, d. h. «ja / ja»-Antworten als in den anderen Ländern.)

Resultate

Besonders viele veränderte Antworten kamen z. B. aus Peru. Meist veränderten sich diese von «nein» zu «ja». Aus der Grafik sieht man sofort auch, dass diese Regel für die Fragen «(b)» und «(g)» nicht gilt.

In Österreich gab es – ausser zur Frage «(d)» – fast keine Wechsel zwischen erster und zweiter Antwort. Auch in Deutschland gab es verhältnismässig wenig Wechsel.

Tafel 33:
Geordnete
Antwortpaare
zu sieben
Ja/Nein-Fragen,
nach Ländern



F.4 Beispiel 2: Antwortpaare zu Ja/Nein-Fragen nach Themen

In der nächsten Grafik sind die Antworten der 111 Personen aus dem vorangegangenen Beispiel 1 nicht nach Ländern, sondern nach den Fragen gebündelt.

Auch hier fallen zuerst die unterschiedlichen Farbanteile auf: Bei Frage «(b)» überwiegen z. B. die hellblauen «nein / nein»-Felder deutlich. Bei den Fragen «(e)» und «(f)» fallen die wenigen blau gefärbten «ja / nein»- und «nein / nein»-Felder auf.

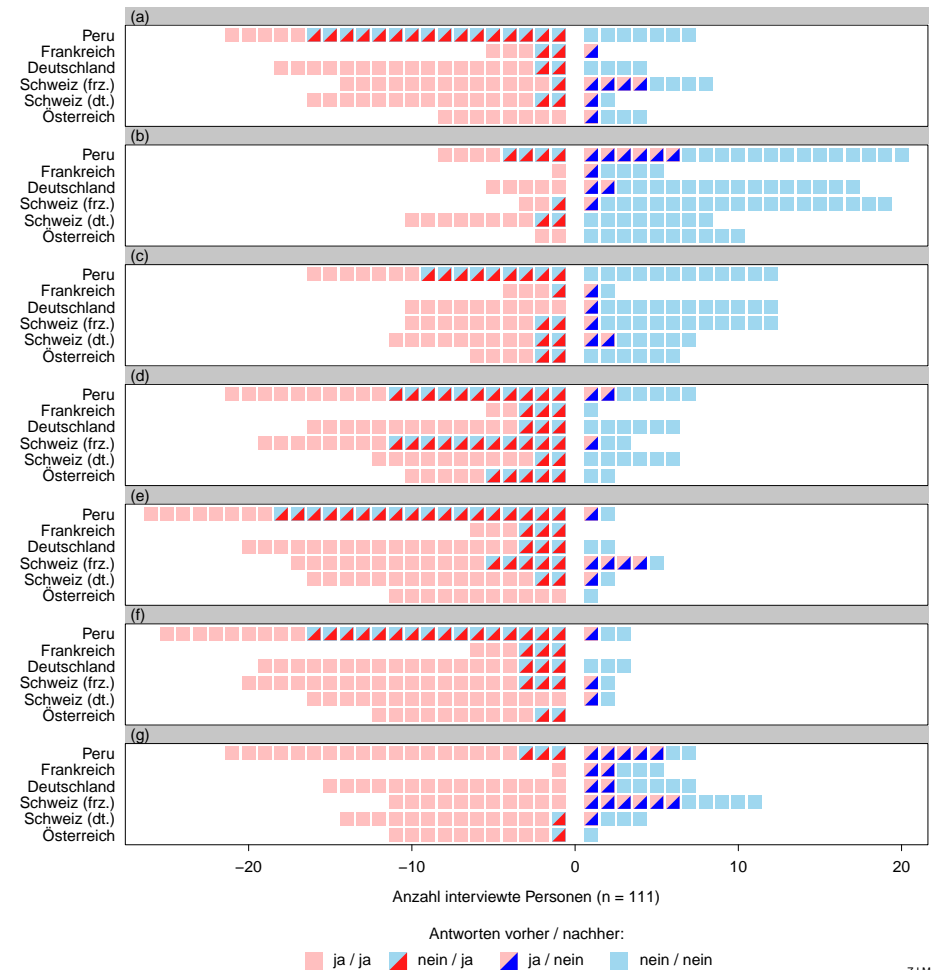
Viele Veränderungen ins Positive (rote Dreiecke: von «nein» zu «ja») ausserhalb Perus gab es bei der Frage «(d)». Sie kamen vor allem aus der französischen Schweiz und aus Österreich. In Deutschland und in der deutschen Schweiz gab es zu dieser Frage hauptsächlich Antworten ohne Wechsel. Am meisten Veränderungen ins Negative (blaue Dreiecke: von «ja» zu «nein») sind bei den Antworten zur Frage «(g)» festzustellen.

Wenn man die Antworten zeilenweise durchschaut, sieht man auch in dieser Grafik, dass das Antwortverhalten in den Ländern unterschiedlich war. Besonders viele Antwortwechsel kamen oftmals aus Peru. Dieser Sachverhalt als Ganzes war allerdings in der vorangegangenen Grafik, die nach Ländern gruppiert war, noch deutlicher zu sehen. Hier nun kann man fragebezogene Details sehen wie z. B., dass zur Frage «(d)» praktisch gleich viele Antwortwechsel wie aus Peru auch aus der französischen Schweiz kamen (bei insgesamt etwas weniger Antworten).

Grafik
► Tafel 34

Resultate

Tafel 34: Geordnete Antwortpaare zu sieben Ja/Nein-Fragen, nach Themen



F.5 Beispiel 3: Antwortkombinationen zu Ja/Nein-Fragen pro Person

In der nächsten Grafik sind alle Antworten pro Person (gruppiert nach Ländern) abgebildet. In den Spalten sieht man, wie die einzelnen Personen geantwortet haben. (In den Zeilen sieht man, wie einzelne Fragen beantwortet wurden.) Die Personen wurden sortiert nach der Anzahl bejahender Antworten.¹ Links wurden jene Personen eingetragen, welche sowohl bei der ersten wie auch bei der zweiten Antwort (meist) mit «ja» geantwortet haben. Ganz rechts sind die Antworten jener Personen zu sehen, die viele «nein»-Antworten abgaben.

Auch in dieser Grafik fällt als Erstes auf, dass sich Antworten aus den verschiedenen Ländern nach der Häufigkeit von Wechseln in der Antwort unterscheiden: Am meisten Dreiecke hat es bei den Antworten aus Peru. Die meisten Wechselflaggen sind rot gefärbt, d. h. sie gingen nach «ja».

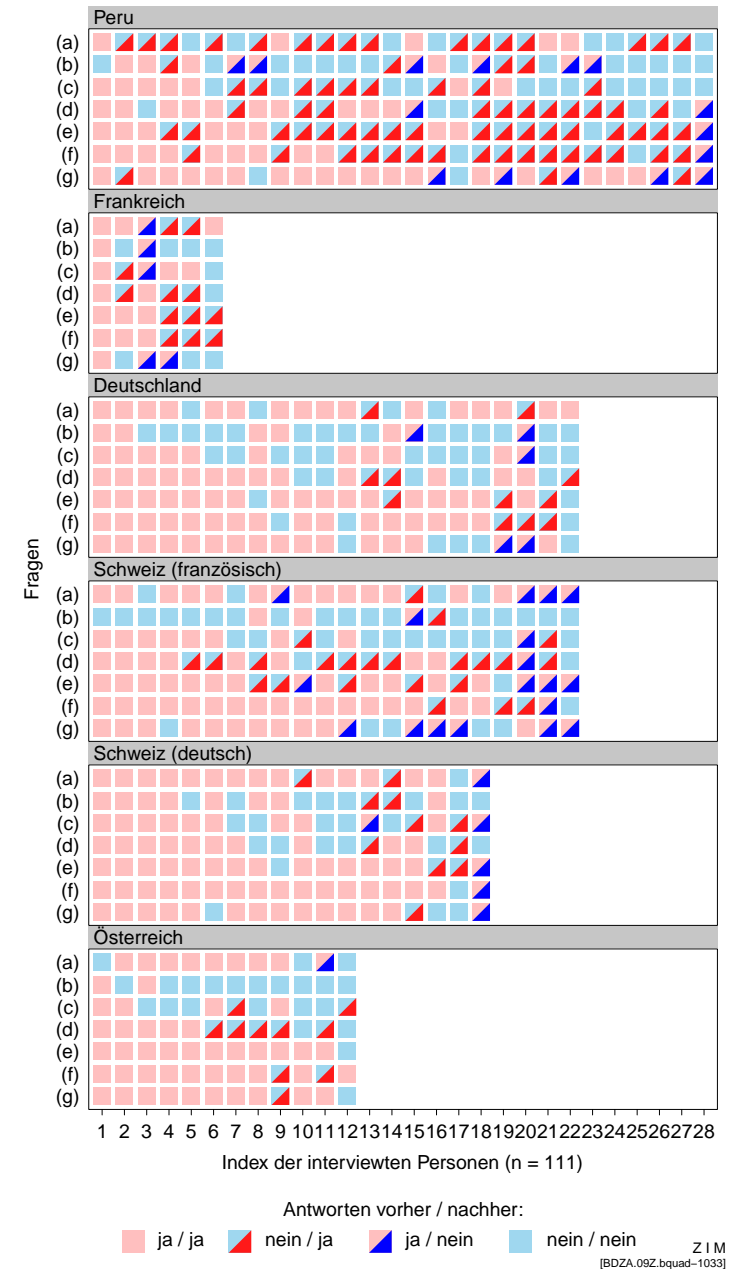
Wenn man die Antworten spaltenweise durchschaut, sieht man u. a. sehr gut, welche Personen beim zweiten Interview lauter Nein-Antworten gaben: Ganz rechts in den Grafikfeldern sind einzelne Spalten zu finden, die nur hellblaue Rechtecke und blaue Dreiecke enthalten. In Peru, in der deutschen und in der französischen Schweiz gab es je eine solche Person. Umgekehrt sind in keinem der Länder ausser in Peru Personen zu finden, welche anfänglich einige Nein-Antworten gaben und im zweiten Interview dann alle Fragen mit Ja beantworteten. Bei den PeruanerInnen führt die Suche zu zwei solchen Spalten ohne hellblaue Felder und ohne Felder mit blauem Dreieck. (Sie befinden sich links an zweiter und an vierter Stelle.)

Grafik
► Tafel 35

Resultate

Tafel 35:

Antwortpaare pro Person zu sieben Ja/Nein-Fragen, nach Ländern und Personen

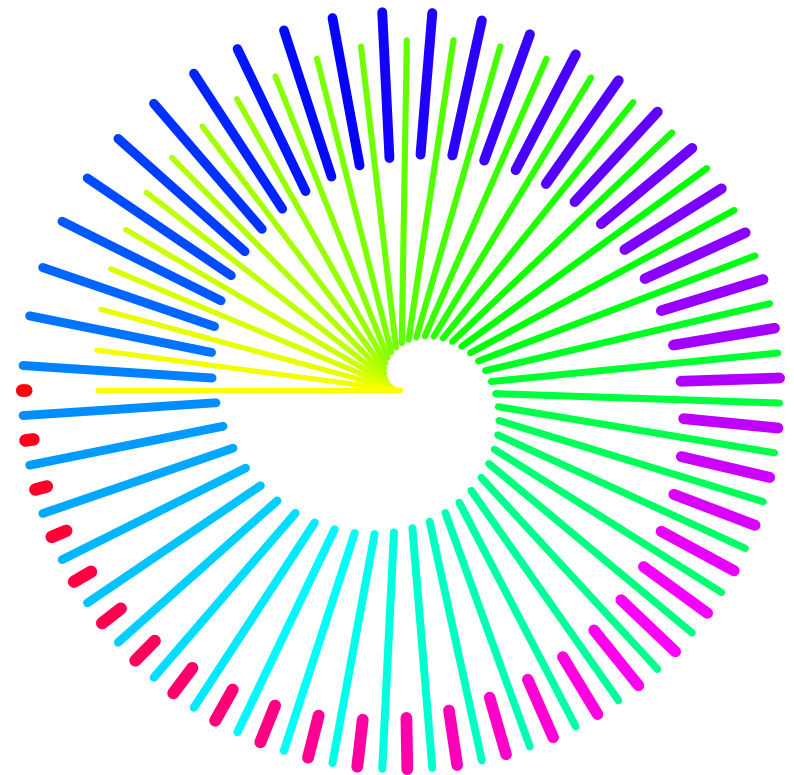


¹ Zum Problem von derartigen Sortierungen vgl. z. B. Bertin [Grafik, 1982]: 32 ff und Hahsler et al. [Seriation, 2008].

Im Anhang folgen einige Anmerkungen zu statistischen Fachausdrücken, welche im Rahmen der Präsentation der Streuungsfächerkarten verwendet wurden. Sie sind für Leser gedacht, denen diese Dinge noch nicht besonders geläufig sind.

↑ S. [31](#)

G Anhang



Speichenspirale 1 (2009)

Speichendifferenzdiagramm mit generierten Daten.

↑ Vgl. [Tafel 3](#) (S. [25](#))

G.1 Statistische Anmerkungen

G.1.1 Median und Quartile in Boxplot und Streuungsfächerkarte

Median

Zur Beschreibung der Lage von Messwerten, die nicht normalverteilt sind, kann der Median (m) verwendet werden. Er gibt denjenigen Messwert an, der «in der Mitte» der Stichprobe liegt: Alle Messwerte werden sortiert; der Wert in der Mitte wird als Median bezeichnet.¹ Vom Medianwert aus gesehen sind 50 % der Messwerte kleiner (oder gleich gross) und 50 % der Messwerte grösser (oder gleich gross).


Bei perfekt normalverteilten Daten ist der Median gleich gross wie das arithmetische Mittel. Es gibt jedoch Verteilungen, die selten normalverteilt sind. Dazu gehören z. B. Verweildauern von Krankenhauspatienten: Es gibt meist einige wenige Behandlungsfälle mit langen und sehr langen Verweildauern und demgegenüber viele mit kürzeren Verweildauern. Bei einer solchen Verteilung liegt der Median unterhalb des arithmetischen Mittels, und man spricht von einer «rechtsschiefen» Verteilung.

Der Median wird – im Unterschied zum arithmetischen Mittel – von Ausreisserwerten nicht beeinflusst. Somit ist er ein *robuster* Mittelwert.

Quartile

Wenn man die beiden Hälften der Stichprobe unter und über dem Median nochmals halbiert, erhält man die Quartile. Quartile unterteilen eine Stichprobe bei 25 % (erstes Quartil), 50 % (zweites Quartil = Median) und 75 % (drittes Quartil). Zwischen dem ersten und dem dritten Quartil liegen also auch gerade 50 % der Messwerte. Dieser Bereich wird «Quartilsabstand» genannt.²

Visualisierung mit Boxplots

Grafisch dargestellt werden Median und Quartile meist in Form von sogenannten «Boxplots». Sie bestehen aus zwei aneinander gefügten Kästchen, welche erstes Quartil, Median und zweites Quartil anzeigen. An jedes der beiden Kästchen ist eine Linie angefügt, welche in der hier verwendeten Form das 5 %- und 95 %-Perzentil anzeigen und somit 90 % der Messwerte umfassen: .³

Aus Boxplots wird auf einen Blick ersichtlich, ob eine schiefe Verteilung vorliegt: In einem solchen Fall sind die beiden Kästchen nicht gleich gross.

Visualisierung mit Streuungsfächern

In Streuungsfächern wird der Median mit der mittleren weissen Linie, die Werte vom ersten bis zum dritten Quartil mit dem schwarzen Fächer, 5 %- und 95 %-Perzentil mit den weissen Federn abgebildet.

↑ S. 32

¹ Bei einer geraden Anzahl von Messwerten wird der Median als Durchschnitt der beiden in der Mitte liegenden Werte berechnet.

² Englisch: IQR = Interquartile Range.

³ Im Boxplot nach Tukey werden die «whiskers» beim $\langle 3. \text{Quartil} + k \times \text{IQR} \rangle$ bzw. beim $\langle 1. \text{Quartil} - k \times \text{IQR} \rangle$ gesetzt (IQR = Interquartilsabstand). Tukey schlug $k = 3.0$ vor; heute wird oft $k = 1.5$ verwendet. – Tukey [EDA, 1977], zitiert in: Nagel et al. [Grafische Datenanalyse, 1996]: 35 ff.

G.1.2 Varianzreduktion (r^2)

Die Varianzreduktion (r^2) ist ein Mass dafür, inwieweit es durch eine Gruppenbildung gelungen ist, die Streuung von (normalverteilten) Daten zu erklären.

Zur Berechnung der Varianzreduktion werden zunächst die Unterschiede der Messwerte x_i zum Mittelwert der Gesamtstichprobe von n Messwerten berechnet, quadriert und dann summiert. (Im grafischen Beispiel beträgt der Mittelwert der Gesamtstichprobe 3.9. Die quadrierten Abweichungen sind in der Zahlenspalte links eingetragen. Die Summe der quadrierten Abweichungen gegenüber dem Mittelwert der Gesamtstichprobe beträgt 19.40.) Dieser Wert für die gesamte Streuung wird verglichen mit der Summe der Werte der Streuungen in Bezug auf die Gruppenmittelwerte. Dazu werden die Unterschiede der Messwerte x_i zu den Gruppenmittelwerten berechnet und ebenfalls quadriert und summiert. (Im Beispiel ergibt dies $5.60 + 8.40 = 14.00$.)

Im nächsten Schritt berechnet man nun den «Unterschied der Gesamtstreuung und der Summe der Streuungen in Bezug auf die Gruppenmittelwerte». (Im Beispiel ist dies $19.40 - 14.00 = 5.40$.) Diese Differenz setzt man anschliessend ins Verhältnis zur Gesamtstreuung. (Im Beispiel ergibt dies $5.40 / 19.40 = 27.8\%$.) Damit erhält man die Varianzreduktion: eine Zahl, welche angibt, um wieviele Prozent die Streuung in Bezug auf die Gruppenmittelwerte kleiner ist als die Streuung in Bezug auf den Mittelwert der Gesamtstreuung.

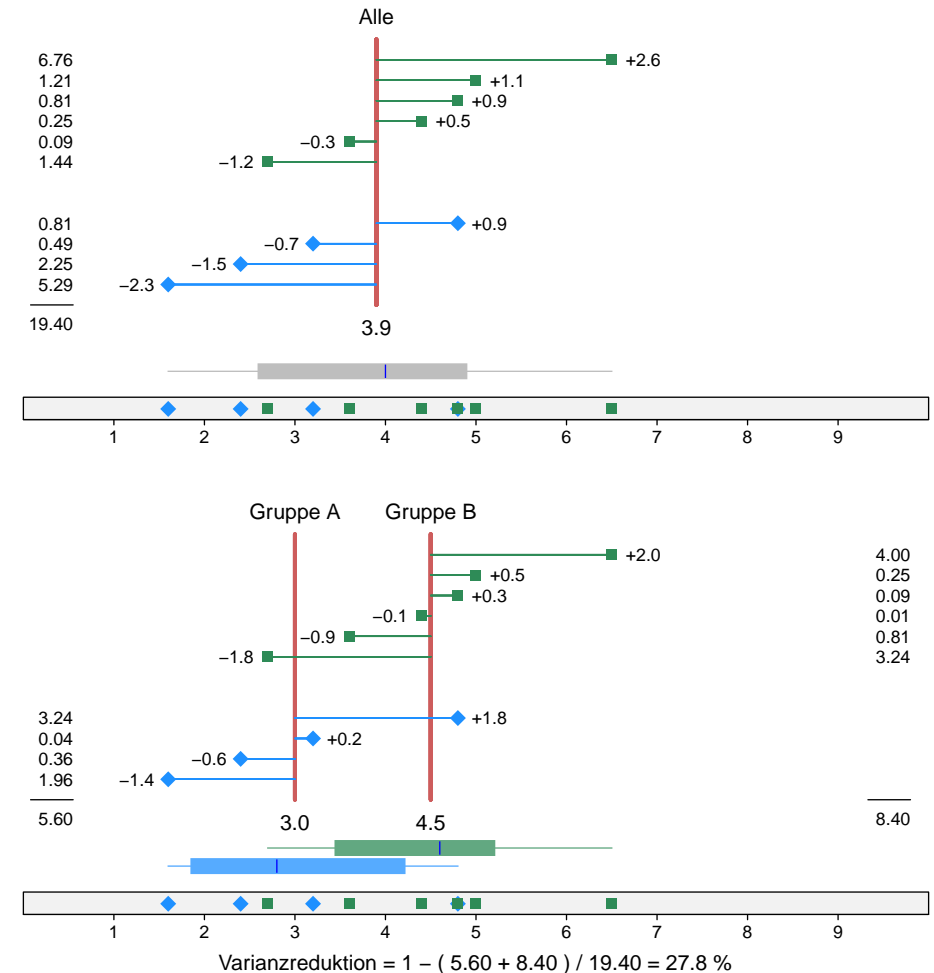
Die – gekürzte – Formel dazu lautet:

$$\text{Varianzreduktion} = 1 - \frac{\sum_{i=1}^n (x_i - \text{Gruppenmittelwert})^2}{\sum_{i=1}^n (x_i - \text{Mittelwert der Gesamtstichprobe})^2}$$

Varianzreduktion (r^2)

► Tafel 36

Tafel 36: Beispiel zur Berechnung der Varianzreduktion



Quelle: Nach Fischer [Grafiken zur PCS-Beurteilung, 2008]: 33

G.1.3 Reduktion der absoluten Abweichungen vom Median (r1)

Alternativ zur Varianzreduktion und als robustere Kennzahl die «Reduktion der absoluten Abweichungen vom Median» ($r1$), um zu beurteilen, inwieweit es durch eine Gruppenbildung gelungen ist, die Streuung von Daten zu erklären. Im Unterschied zur Varianzreduktion wird sie viel weniger stark von Extremwerten beeinflusst, denn die Abweichungen zu den Gruppenmittelwerten werden nicht quadriert, sondern nur als absolute Werte aufsummiert.

Im grafischen Beispiel wurden die durchschnittlichen absoluten Abweichungen vom Median in den Zahlenspalten links und rechts aussen unter den Summen der absoluten Abweichungen eingetragen. Sie betragen in den beiden Gruppen 1.00 und 0.93. Wenn keine Gruppierung vorgenommen wird, liegt dieser Wert bei 1.20; er ist also grösser als beide Gruppenwerte.

Reduktion der absoluten Abweichungen vom Median ($r1$)

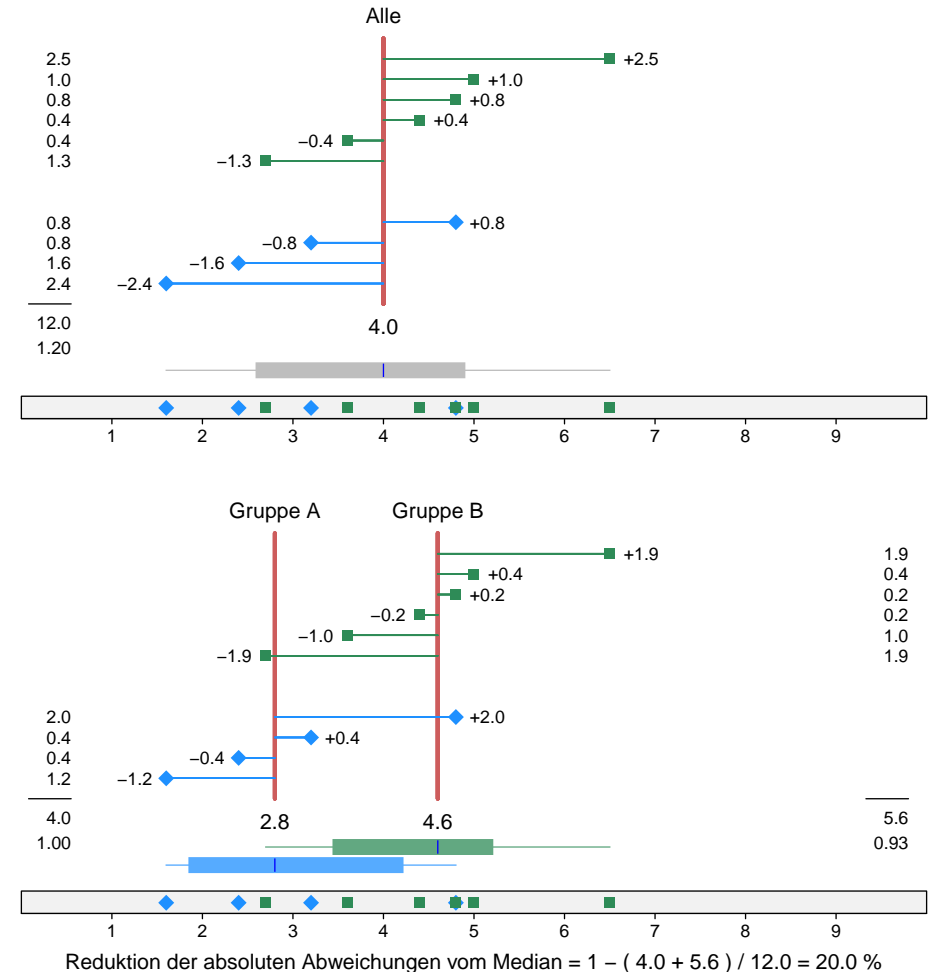
► Tafel 37

$$\text{Red. d. abs. Abw. v. Med.} = 1 - \frac{\sum_{i=1}^n |x_i - \text{Gruppenmedian}|}{\sum_{i=1}^n |x_i - \text{Median der Gesamtstichprobe}|}$$

Um diese Formel besser zu verstehen, kann man auch zuerst die Differenz zwischen der «Summe der Abweichungen vom Gruppenmedian» und der «Summe der Abweichungen vom Gesamtmedian» berechnen. Anschliessend wird das Resultat dieser Berechnung durch die «Summe der Abweichungen vom Gesamtmedian» dividiert. Damit hat man berechnet, wie gross die Reduktion der summierten Abweichungen von den mittleren Werten bei Gruppenbildung im Verhältnis zu den Abweichungen vom mittleren Wert ohne Gruppenbildung ist.

(Es kann noch angemerkt werden, dass grundsätzlich gilt: Die Summe der absoluten Abweichungen von einem Wert aus einer Messreihe ist dann am kleinsten, wenn es sich bei diesem Wert um den Median handelt.)

Tafel 37: Beispiel zur Berechnung der Reduktion der absoluten Abweichungen vom Median



Quelle: Nach Fischer [Grafiken zur PCS-Beurteilung, 2008]: 34

G.2 Verzeichnisse

G.2.1 Abkürzungen

Tafel 38: Im Text verwendete Abkürzungen

Abkürzung	Bezeichnung	Übersetzung; Internetverweise
\bar{x}	Durchschnitt (arithmetisches Mittel)	
a	Durchschnittliche absolute Abweichung vom Median	
APDRG	All Patient Diagnosis Related Groups	http:// www.fischer-zim.ch / texttk-pcs / t-B3-drg-fam-AP-0801.htm
BAG	Bundesamt für Gesundheit	http:// www.bag.admin.ch /
BFS	Bundesamt für Statistik	http:// www.bfs.admin.ch /
DRG	Diagnosis Related Groups	http:// www.fischer-zim.ch / texttk-pcs / index.htm
FIM	Functional Independence Measure	http:// www.udsmr.org /
HRG	Healthcare Resource Groups	http:// www.ic.nhs.uk / our-services / standards-and-classifications / casemix
IQR	Interquartile Range	Quartilsabstand
KVG	Krankenversicherungsgesetz vom 18.3.1994	http:// www.admin.ch / ch / d / sr / c832_10.html
LEP	Leistungserfassung in der Pflege	http:// www.lep.ch /
m	Median	
OKP	Obligatorische Krankenpflege-Grundversicherung	http:// www.praemien.admin.ch /
PCS	Patientenklassifikationssystem	http:// www.fischer-zim.ch / studien / PCS-Buch-9701-Info.htm
$r1$	Reduktion der absoluten Abweichungen vom Median	↑ S. 102
$r2$	Varianzreduktion	↑ S. 100
s	Standardabweichung	
SGP	Schweizerische Gesellschaft für Pädiatrie	http:// www.swiss-paediatrics.org /
SZH	Stiftung Zürcher Höhenkliniken	http:// www.zhwh.ch / 031dav _ 0103 _ de.htm
TAR	Leistungsbedarfsbezogenes Tarifsystm für Rehabilitationskliniken	http:// www.fischer-zim.ch / studien / TAR-RehaPCS-0607-Info.htm

G.2.2 Literaturverzeichnis

Becker/Cleveland (1996) Trellis/Man

Becker RA, Cleveland WS. *Trellis Graphics User's Manual*. Seattle (MathSoft, Inc.), Murray Hill (Bell Labs) 1996: 205 S. Internet: [http:// cm.bell-labs.com / stat / doc / trellis.user.pdf](http://cm.bell-labs.com/stat/doc/trellis.user.pdf).

Becker et al. (1996) Trellis

Becker RA, Cleveland WS, Shyu MJ. *The Visual Design and Control of Trellis Display*. In: Journal of Computational and Statistical Graphics 1996/5: 123–155. Internet: [http:// cm.bell-labs.com / stat / doc / trellis.jcgs.col.ps](http://cm.bell-labs.com/stat/doc/trellis.jcgs.col.ps).

Bertin (1982) Grafik

Bertin J. *Graphische Darstellungen und die graphische Weiterverarbeitung der Information*. Berlin New York (Gruyter) 1982 [1977]: 275 S.

Brügger et al. (2002) LEP-Methode 2.0

Brügger U, Bamert U, Maeder C, Odermatt R. *Beschreibung der Methode LEP® Nursing 2*. Leistungserfassung für die Gesundheits- und Krankenpflege. 2., überarbeitete Auflage, (LEP AG) 2002: 32 S. Internet: [http:// www.lep.ch / pdf / Methodenbeschreibung_2_1.pdf](http://www.lep.ch/pdf/Methodenbeschreibung_2_1.pdf).

Chambers (1998) S

Chambers JM. *Programming with Data*. A Guide to the S Language. New York (Springer) 1998. Info: [http:// cm.bell-labs.com / cm / ms / departments / sia / Sbook](http://cm.bell-labs.com/cm/ms/departments/sia/Sbook).

Chambers (2008) R

Chambers JM. *Software for Data Analysis*. Programming with R. New York (Springer) 2008: 498 S.

Cleveland (1993) Visualizing

Cleveland WS. *Visualizing Data*. Summit (Hobart) 1993: 360 S.

Dalgaard (2002) R

Dalgaard P. *Introductory Statistics with R*. New York (Springer) 2002: 267 S.

DH-UK (2007) PbR 2008/09

Dept. of Health Payment by Results Finance and Costing Team. *Tariff Information: Confirmation of Payment by Results (PbR) arrangements for 2008/09*. Exceltable, 2007. Download ab: [http:// www.dh.gov.uk / en / Managingyourorganisation / Financeandplanning / NHSFinancialReforms / DH _ 081226](http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/DH_081226).

Fetter et al. (1991) DRGs

Fetter RB, Brand A, Dianne G [Hrsg.]. *DRGs, Their Design and Development*. Ann Arbor (Health Administration Press) 1991: 341 S.

Fischer (1997) PCS

Fischer W. *Patientenklassifikationssysteme zur Bildung von Behandlungsfallgruppen im stationären Bereich*. Prinzipien und Beispiele. Bern und Wolfertswil (ZIM) 1997: 514 S. Auszüge: [http:// www.fischer-zim.ch / studien / PCS-Buch-9701-Info.htm](http://www.fischer-zim.ch/studien/PCS-Buch-9701-Info.htm).

Fischer (2002) DRG+Pflege

Fischer W. *Diagnosis Related Groups (DRGs) und Pflege*. Grundlagen, Codierungssysteme, Integrationsmöglichkeiten. Bern (Huber) 2002: 472 S. Auszüge: [http:// www.fischer-zim.ch / studien / DRG-Pflege-0112-Info.htm](http://www.fischer-zim.ch/studien/DRG-Pflege-0112-Info.htm).

Fischer (2005) Krankenhaus-Betriebsvergleiche

Fischer W. *Neue Methoden für Krankenhaus-Betriebsvergleiche*. Ein Werkstattbuch zur Visualisierung DRG-basierter Daten. Wolfertswil (ZIM) 2005: 160 S. Auszüge: [http:// www.fischer-zim.ch / studien / KBV-0506-Info.htm](http://www.fischer-zim.ch/studien/KBV-0506-Info.htm).

Fischer et al. (2006) TAR und Reha-PCS

Fischer W, Blanco J, Mäder M, Zangger P, Conti FM, et al.. *Das TAR-System und andere Patientenklassifikationssysteme für die Rehabilitation*. TAR-Forschungsbericht und Kurzbeschreibung von Systemen aus Deutschland, Frankreich, Australien und den USA. Wolfertswil (ZIM) 2006: 89 S. Info: <http://www.fischer-zim.ch/studien/TAR-RehaPCS-0607-Info.htm>.

Fischer (2008) Grafiken zur PCS-Beurteilung

Fischer W. *Statistische Grafiken zur Beurteilung von Patientenklassifikationssystemen*. dargestellt am Beispiel der pädiatrischen Sicht auf das APDRG-System. Wolfertswil (ZIM) 2008: 169 S. Internet: <http://www.fischer-zim.ch/studien/Grafiken-PCS-Beurteilung-0804.htm>.

Fischer (2008) DRG-Familie

Fischer W. *Die DRG-Familie*. Stand: 2007. Wolfertswil (ZIM) 2008: 32 S. Internet: <http://www.fischer-zim.ch/textk-pcs/index.htm>.

Fischer (2009) Notfallvergütung im Krankenhaus

Fischer W. *Notfallvergütung im Krankenhaus*. Patientenklassifikationssysteme und Notfallpauschalen bei DRG-basierter Vergütung von stationären Behandlungen. 1. Auflage, Wolfertswil (ZIM) 2009: 180 S.

Granger et al. (1993) FIM-Rasch

Granger CV, Hamilton BB, Linacre JM, Heinemann AW, Wright BD. *Performance Profiles of the Functional Independence Measure*. In: Am J Phys Med Rehabil 1993(72): 84–89.

Granger et al. (1995) FIM

Granger CV, Brownschidle CM. *Outcome Measurement in Medical Rehabilitation*. In: International Journal of Technology Assessment in Health Care 1995(11)2: 262–268.

Granger et al. (1995) UDSmr-93

Granger CV, Ottenbacher KJ, Fiedler RC. *The Uniform Data System for Medical Rehabilitation*. In: American Journal of Physical Medicine & Rehabilitation 1995(74)1: 62–66.

Hahsler et al. (2008) Seriation

Hahsler M, Hornik K, Buchta C. *Getting Things in Order: An Introduction to the R Package seriation*. In: Journal of Statistical Software 2008(25)3. Internet: <http://www.jstatsoft.org/v25/i03/>.

Nagel et al. (1996) Grafische Datenanalyse

Nagel M, Benner A, Ostermann R, Keschke K. *Grafische Datenanalyse*. Stuttgart Jena New York (Gustav Fischer) 1996: 292 S.

Tufte (2001) Display

Tufte ER. *The Visual Display of Quantitative Information*. Second Edition, Cheshire (Graphics Press) 2001 [1983]: 200 S.

Tukey (1977) EDA

Tukey JW. *Exploratory Data Analysis*. Reading (Addison Wesley) 1977.

UDSMr (1997) FIM™ instrument

UDSMr. *Guide for the Uniform Data Set for Medical Rehabilitation (including the FIM™ instrument)*. Version 5.1. Buffalo (University at Buffalo) 1997.

Unwin et al. (2006) Graphics

Unwin A, Theus M Hofmann H. *Graphics of Large Datasets*. Visualizing a Million. Berlin New York (Springer) 2006: 271 S.

Stichwortverzeichnis

Balken mit Dreieck	56	in der Speichengrafik	19
Balkendiagramm	► Punktbalkendiagramm, 20, 54	Kunstgrafik	11
Barley	► Gerstenernte	LaTeX	4
Betrachtungsdistanz	13, 40	LEP = Leistungserfassung in der Pflege	34, 36
Boxplot	32, 99	Lesbarkeit	13, 78
Differenzdiagramm	47, 54	Liniendiagramm	73, 80
Doppeldifferenzdiagramm	56	Lupe	12, 22
Kombination von D. diagrammen	58	Markierungsstrich	
Distanz	► Betrachtungsdistanz	bei Unsichtbarkeit	19
Dreieck	47	Mathematikprüfungen	38
im Sequenzdiagramm	75	Median	31, 99, 102
mit Balken	56	im Boxplot	99
DRG = Diagnosis Related Groups	► HRG, 21, 48	im Streuungsfächer	32, 99
Fächer	► Streuungsfächerkarte	Pdf-Datei	
Fachwerkgrafik	36, 50	aktive Querverweise	12
Fallgewichte		Pseudogeografische Anordnung	40
HRG: elektiv / nicht-elektiv	24	Punktbalkendiagramm	52, 54
Fan chart	► Streuungsfächerkarte	Punktendiagramm	54
Farben	14	Quartil	99
Federn		Boxplot	99
im Streuungsfächer	32	Streuungsfächer	32
FIM = Functional Independence Measure	34, 36	Quartilsabstand	99
Streuungsfächer	32	Streuungsfächer	32
Gerstenernte		Reduktion der absoluten Abweichungen	
Minnesota 1931 und 1932	50	vom Median	64, 102
Grafiktypen		Rehabilitation	34, 36, 76, 78, 80
bquad	87	Relevanz	13, 31
diffplot	47	Relevanzbalken	36
dotbar	52	Sequenzbänder	78
fanchart	31	Sequenzdiagramm	75
seqplot	73, 78, 80	Speichengrafik	19, 20
spokeplot	19	Spirale	96
tri2x	58	Spoke plot	► Speichengrafik
tribar	56	Strömungsbild	31, 66
Histogramm	19	Streuungsfächer	32
HRG = Healthcare Resource Groups	24	Streuungsfächerkarte	31
Interquartilsabstand	► Quartilsabstand	Trellis display	► Fachwerkgrafik
Kammlinie	80	Varianzreduktion	64, 100
Krankenkassenprämien	40	Wechseldiagramm	87
Kreis		Wechselflagge	88
im Streuungsfächer	32	Windrichtung	31

Wolfram Fischer

Neue Methoden für Krankenhaus-Betriebsvergleiche

Ein Werkstattbuch zur Visualisierung DRG-basierter Daten

2005, 160 S., 109 Tab. und meist farbige Abb., geb.

€ 25.90 / SFr. 38.80

ISBN 978-3-9521232-8-7 (ZIM)

Gleich zu Beginn ist in diesem Buch ein umfassendes Beispiel eines Krankenhaus-Betriebsvergleichs zu finden. Anschliessend werden Schritt für Schritt Lösungen zu den folgenden Fragen präsentiert: 1. Wie lässt sich ein Krankenhaus beschreiben? 2. Wie können ähnliche Krankenhäuser ermittelt werden? 3. Welche Kennzahlen und Grafiken ermöglichen diskussionsfähige Krankenhausvergleiche? Dazu werden neuentwickelte «dichte» Grafiken verwendet.



Wolfram Fischer

Statistische Grafiken zur Beurteilung von Patientenklassifikationssystemen

dargestellt am Beispiel der pädiatrischen Sicht auf das APDRG-System

2008, 169 S., 81 Tab. und meist farbige Abb., kt.

€ 29.90 / SFr. 45.00

ISBN 978-3-905764-03-1 (ZIM)

Auswertungen von DRG-Daten tendieren zur Unübersichtlichkeit. In diesem Buch wird gezeigt, wie es mit «dichten» Grafiken möglich wird, einen Überblick zu schaffen, und wie gleichzeitig Problembe-
reiche sichtbar gemacht werden können.

