

Die Software



Projekt tinkerplots.de

Link zur Projektseite:

<http://lama.uni-paderborn.de/personen/rolf-biehler/projekte/tinkerplots.html>

Kontakt

Rolf Biehler: biehler@math.upb.de (Telefon: 05251 – 602654)

Daniel Frischemeier: dafr@math.upb.de (Telefon: 05251 – 603229)

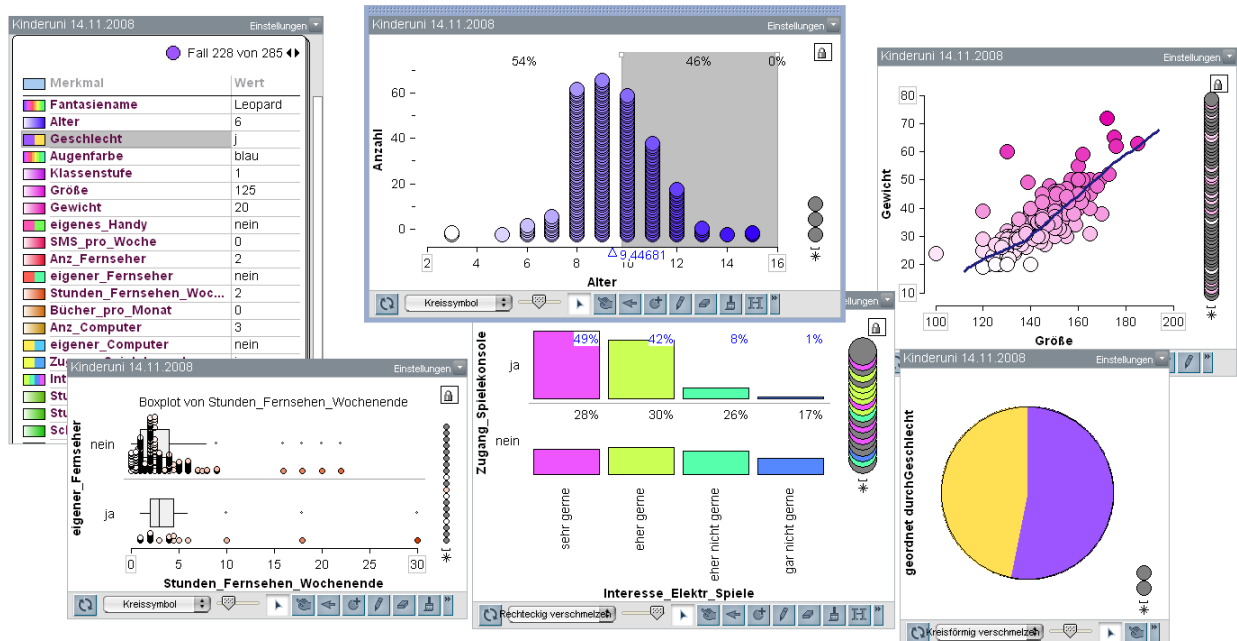
Susanne Podworny: podworny@math.upb.de (Telefon: 05251 – 603229)

Inhaltsverzeichnis

1. Datenanalyse mit der Software TinkerPlots	4
1.1 Grundlegendes	4
1.1.1 Der TinkerPlots Bildschirm	4
1.1.2 Die Menüleiste	4
1.2 Motivation: Statistische Auswertungen mit Datenkarten	6
1.3 Einfache Auswertungen mit TinkerPlots.....	9
1.4 Erste Verteilungsvergleiche in TinkerPlots	14
2. Stochastische Simulationen mit der Zufallsmaschine in TinkerPlots	27
2.1 Die Zufallsmaschine in der Übersicht	27
2.2 Die einzelnen Bauteile der Zufallsmaschine.....	28
3. Ausgewählte Aufgaben zur	30
3.1 Datenanalyse mit TinkerPlots	30
3.2 Simulation mit TinkerPlots.....	30
3.2.1 Modellierung des „doppelten Würfelwurfs“ mit der Zufallsmaschine.....	31
3.2.2 Die Tabelle und ihre Formeln am Beispiel des „doppelten Würfelwurfs“	32

Eine Einführung in die elementare Datenanalyse mit der

Software

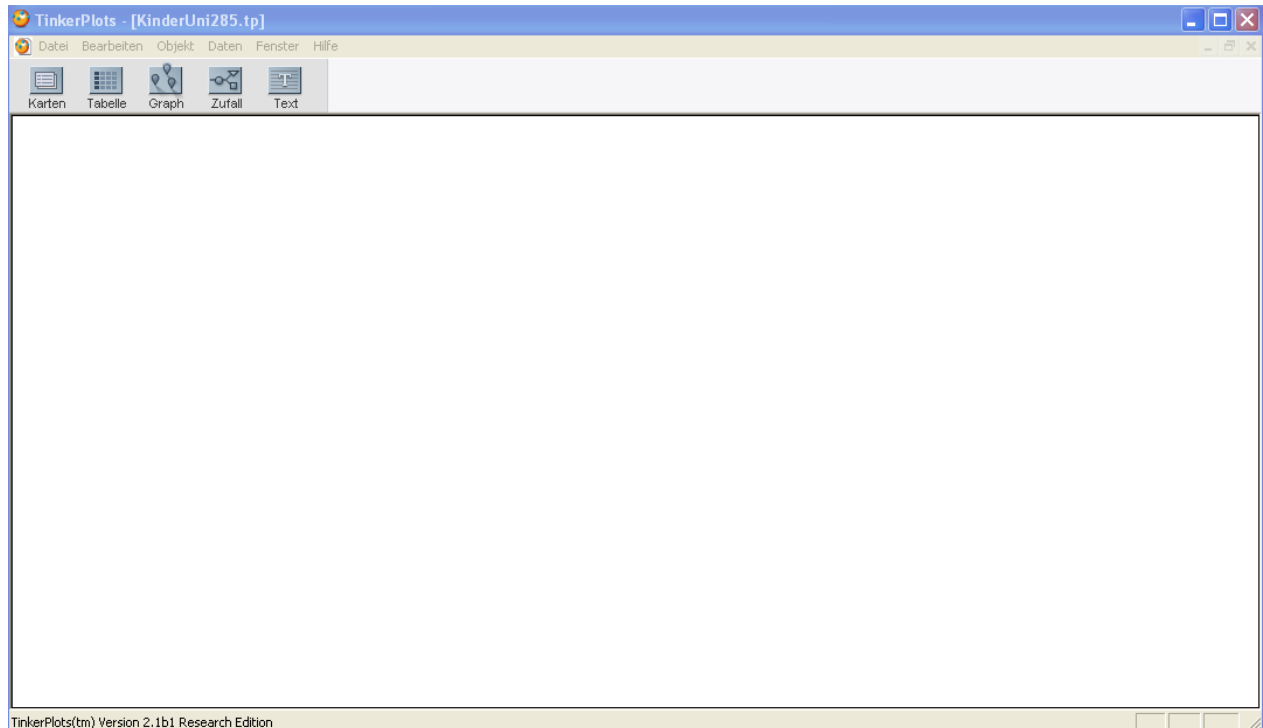


1. Datenanalyse mit der Software TinkerPlots

1.1 Grundlegendes

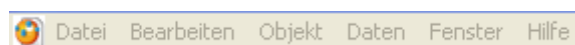
1.1.1 Der TinkerPlots Bildschirm

Wenn Sie das Programm TinkerPlots starten, so finden Sie folgenden Bildschirm vor:



1.1.2 Die Menüleiste

Ähnlich, wie Sie es in zum Beispiel in Windows gewohnt sind, gibt es eine Menüleiste, die die wichtigsten grundlegenden Funktionen, wie Datei öffnen, Datei speichern, Rückgängig-Funktion, Hilfemenü, usw. ... beinhaltet:



Ebenfalls gibt es das Arbeitsfeld und die Hauptleiste, in der Sie Datenkarten, Datentabellen, Graphen, Zufallsmaschinen und Textfelder in das Arbeitsfeld (weiße Fläche) ziehen können.

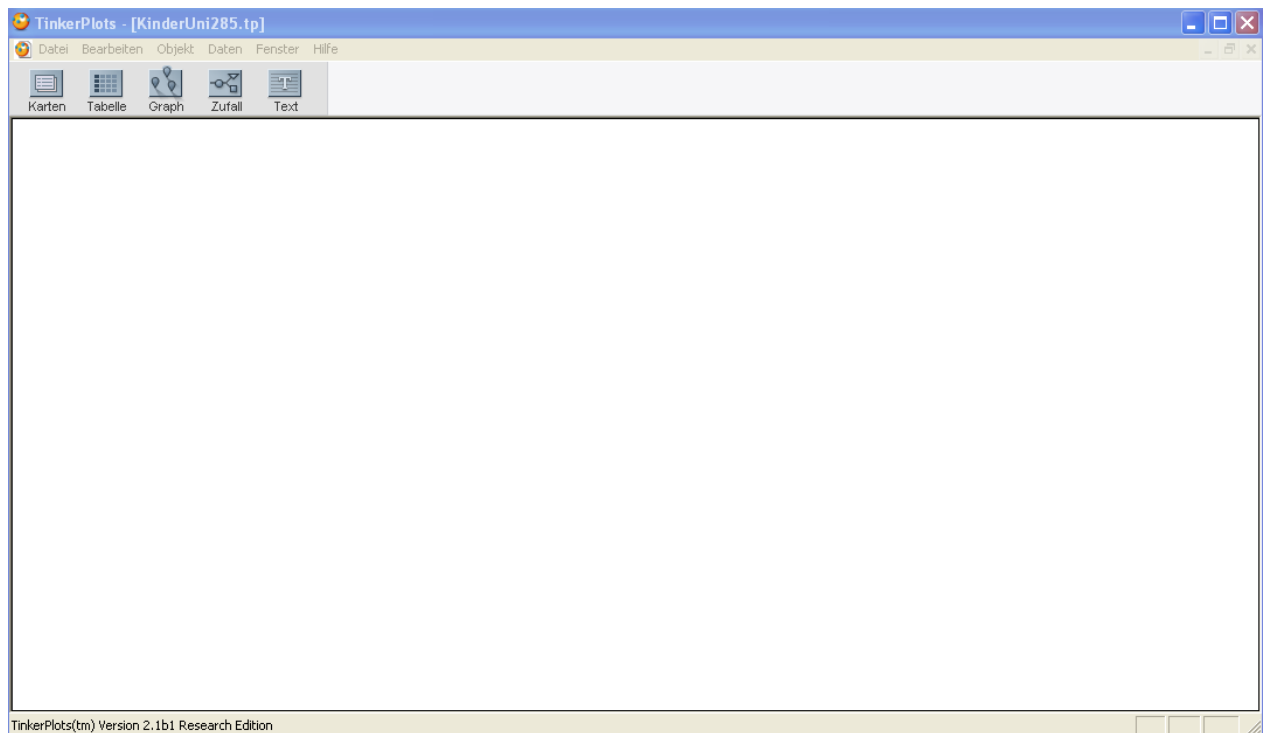
In dieser Einführung wollen wir vor allem auf die Funktionen der

Datenkarten

Datentabelle

Graphen

eingehen. Zunächst ein kurzer zusammenfassender Überblick über die wichtigsten Icons. Die ausführliche Beschreibung folgt in den nächsten Abschnitten.



Karten Dies ist ein Datenkartenstapel. In ihm werden alle Fälle abgelegt. Jeder Fall bekommt seine eigene Datenkarte. Die verschiedenen Merkmale stehen untereinander und sind auf jeder Datenkarte gleich.



Tabelle Alle Daten können auch in einer Tabelle angezeigt werden. Ein Fall befindet sich dann in einer Zeile, eine Zeile entspricht also einer Datenkarte des Datenkartenstapels. Die Merkmale der einzelnen Fälle stehen in den Spalten. Änderungen in der Tabelle führen automatisch zu Änderungen in den Datenkarten und umgekehrt.



Graph Hiermit werden Graphiken aus den Daten erzeugt. Dies ist das Hauptauswertungsobjekt in TinkerPlots.



Zufall Hinter diesem Objekt verbirgt sich die Zufallsmaschine von TinkerPlots. Diese wird für alle Simulationen benötigt. Eine Tabelle mit den erzeugten Daten wird automatisch erstellt, sobald die Zufallsmaschine gestartet wird.



Text Für Erklärungen, Ergebnisse, Hinweise oder sonstigen Text stehen Textfelder zur Verfügung.

1.2 Motivation: Statistische Auswertungen mit Datenkarten

Die Einführung in die Software soll mittels des Beispieldatensatzes „KinderUni285.tp“ vollzogen werden.

Man stelle sich folgende Situation vor: Die Klasse 5c möchte sich am Beginn des Schuljahres besser kennenlernen und beschließt eine kurze klasseninterne Umfrage durchzuführen. Schnell haben Tim, Tom, Tine und Tina Ideen: Wie viele essen gerne Nudeln in unserer Klasse? Wie viele von uns haben ein Haustier? Schreiben die Jungen am Tag mehr SMS als die Mädchen? An dieser Stelle könnte man die Durchführung eines klasseninternen Projektes unter dem Motto „Wir lernen uns besser kennen“ motivieren. Dieses würde neben dem Aufwerfen von „echten“ statistischen Fragestellungen auch das Erstellen einer geeigneten Erhebungsform (meistens Fragebogen) vorsehen. Hat man dieses nun gemacht (z.B. in Form einer Doppelseite beim Freundschaftsbuch) und die Daten erhoben, stellt sich die Frage der Auswertung. Ein geläufiges Hilfsmittel, was im Erfahrungsbereich der Schülerinnen und Schüler liegt, ist die Stichliste. Diese hat allerdings den Nachteil, dass sie eigentlich nur „eindimensionale“ Schlüsse, d.h. Aussagen über die Verteilung eines Merkmals isoliert zulässt (und streng genommen kann man in diesem Falle gar nicht von einer Verteilung im eigentlichen Sinne sprechen, da z.B. die Form einer Verteilung bei einer Strichliste gar nicht direkt ersichtlich wird). Ein anderes, alternatives Vorgehen, was sich anbietet, ist das Arbeiten mit sogenannten Datenkarten: Um ein einfaches einführendes Beispiel zu betrachten, stellen wir uns dieses vor: Jede Schülerin und jeder Schüler der Klasse 5c bekommt ein PostIt-Zettel. Darauf schreibt sie/er einen Fantasienamen (um die Anonymität zu gewährleisten), sein Geschlecht, seine Augenfarbe und z.B. seine Körpergröße (evtl. geschätzt).¹ Nun kann man zunächst einfache „eindimensionale“ Fragestellungen mit unseren Datenkarten beantworten. Wir legen die Datenkarten aller Schülerinnen und Schüler, sagen wir $n=26$, in die Mitte, und versuchen eine Struktur dort hinein zu bekommen. Eine erste Struktur erhalten wir durch das „Trennen“ von Datenkarten, z.B. nach dem Merkmal Geschlecht. Wir haben dann auf der einen Seite die Datenkarten von den Jungen, auf der anderen Seite die Datenkarten von den Mädchen und können uns durch Abzählen oder (nächster Ordnungsdurchgang) Stapeln der Datenkarten einen Überblick über die Verteilung des Merkmals Geschlecht in der Klasse erlauben (um die Klasse auch selbst noch zu aktivieren, können man sich vorstellen, dass man natürlich auch die Schülerinnen und Schüler selbst im Klassenraum trennen könnte (Untergruppen bilden), um das Problem zu veranschaulichen). Nebenbei bemerkt haben wir dann mithilfe unserer beiden „Stapel“ schon eine Art Säulendiagramm erstellt. Nun hätten wir das unter Umständen auch noch mithilfe der Strichliste hinbekommen. Lassen wir unsere Fragestellung aber komplexer werden, so stellt sich das Arbeiten mit Datenkarten als vorteilhaft heraus, z.B. könnten wir fragen: Sind in dieser Klasse die Jungen tendenziell größer als die Mädchen? (und weiterführend: inwiefern unterscheiden sich die Jungen und Mädchen in dieser Klasse hinsichtlich ihrer Körpergröße?). In der folgenden Abbildung sehen wir einige Datenkarten, auf denen (reduziert) nur die Ausprägungen der Merkmale Geschlecht und Körpergröße zu finden sind:

¹ Man kann natürlich auch ein anderes quantitatives Merkmal, wie Zeit_Comp (Wie viele Stunden am Tag verbringst du durchschnittlich am PC?), o.ä. nehmen.



Diese lassen sich jetzt nun z.B. der Größe nach aufsteigend ordnen.



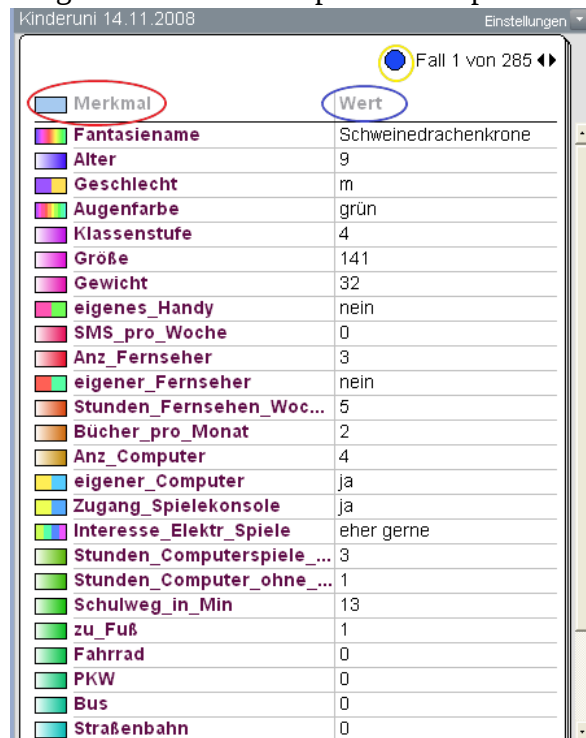
“Trennt” man nun die Schüler von den Schülerinnen, so lassen sich schon erste qualitative Aussagen, wie „die Jungen sind tendenziell größer als die Mädchen“ formulieren.



Das ganze kann man sich natürlich auch mit anderen Merkmalen vorstellen. Die einzigen Operationen, die man benötigt, sind im Wesentlichen das „Ordnen“, „Trennen“ und „Sortieren“ der Datenkarten. Was ist aber nun, wenn wir mehrere Merkmale, sagen wir 10,

und mehrere Schülerinnen und Schüler befragen wollen, sagen wir unsere ganze Schule? Wie können wir die Daten dann effektiv auswerten? Natürlich kann man genauso vorgehen, wie es oben beschrieben wurde, jedoch wäre der Sortierungs- oder auch Trennprozess beispielsweise sehr langwierig. An dieser Stelle kann es nützlich sein, sich von einer Computersoftware unterstützen zu lassen.

Wir betrachten nun die Umsetzung in TinkerPlots: Öffnen Sie den Beispieldatensatz „KinderUni285.tp“. Dieser wurde im Rahmen der Kasseler KinderUni² im Jahr 2008 erhoben und enthält die Angaben von 285 Kindern im Alter von 3 bis 15 zu ihren Freizeitaktivitäten. Diese Daten sind in TinkerPlots, wie in obigen Beispiel, als Datenkarten („Datacards“) - gesammelt. Auf der Abb. links sehen wir exemplarisch eine solche Datenkarte. In der Spalte Merkmal (roter Kreis) können die entsprechenden Merkmale, in der Spalte Wert (blauer Kreis) die zugehörigen Ausprägungen eingetragen werden. Außerdem können numerische Merkmale in der Spalte Einheit (hier nicht zu sehen) mit Einheiten versehen werden. Oben rechts (gelber Kreis) kann man zwischen den verschiedenen Fällen (hier in diesem Falle sind das natürlich die verschiedenen Kinder) hin- und herspringen.



Merkmal	Wert
Fantasiename	Schweinedrachenkrone
Alter	9
Geschlecht	m
Augenfarbe	grün
Klassenstufe	4
Größe	141
Gewicht	32
eigenes_Handy	nein
SMS_pro_Woche	0
Anz_Fernseher	3
eigener_Fernseher	nein
Stunden_Fernsehen_Woc...	5
Bücher_pro_Monat	2
Anz_Computer	4
eigener_Computer	ja
Zugang_Spielekonsole	ja
Interesse_Elekt_Spiele	eher gerne
Stunden_Computerspiele ...	3
Stunden_Computer_ohne ...	1
Schulweg_in_Min	13
zu_Fuß	1
Fahrrad	0
PKW	0
Bus	0
Straßenbahn	0

² Nähere Informationen unter: www.die-kinder-uni.de/html/kassarchiv.html (aufgerufen am 13.7.2012)

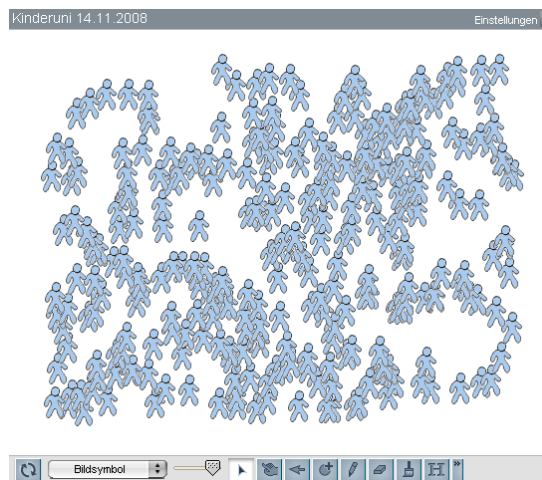
1.3 Einfache Auswertungen mit TinkerPlots

Ein wesentliches Element der Software zum Erforschen von Daten ist die Darstellung im Graphen. Wie schon oben erläutert, ermöglicht er eine Auswertung der Daten in Form von Sortieren (Ordnen, Trennen und Stapeln) der Datenkarten.

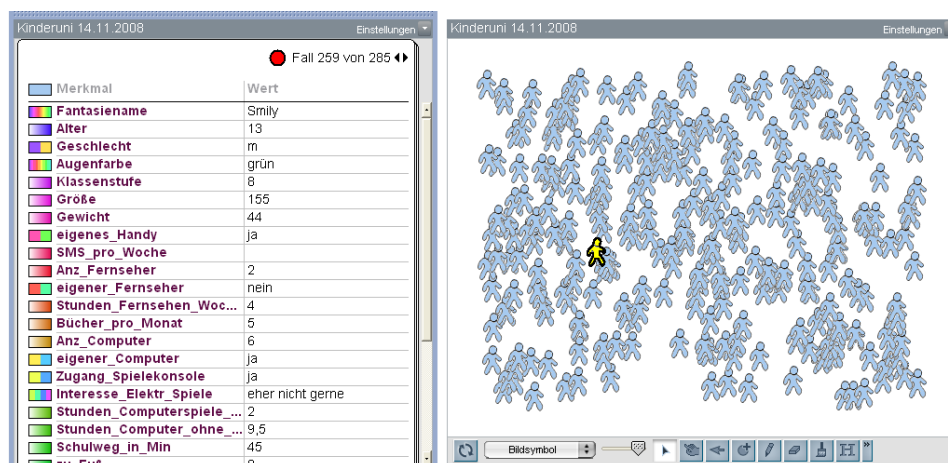
Um den Graph zu einem Datenkartensatz zu erstellen, wählt man in der Arbeitsleiste das Icon



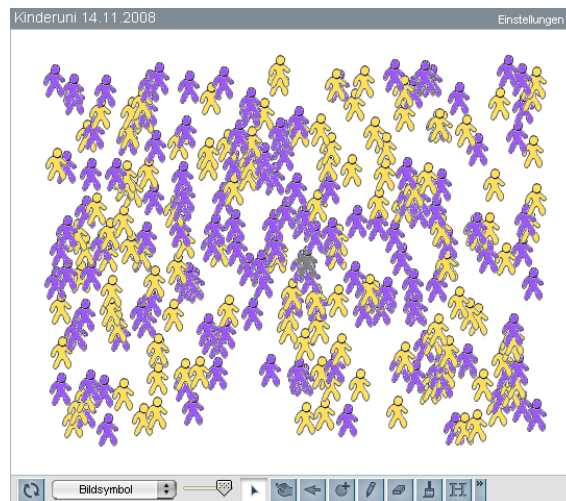
und zieht dieses per „Drag & Drop“ in den Arbeitsbereich. Es öffnet sich ein Fenster mit verschiedenen Punkten (anstelle von Punkte haben wir, ähnlich, wie wir es von Piktogrammen kennen, Figuren):



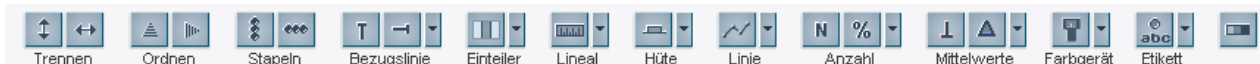
Jedes Symbol im Graph steht für ein bestimmtes Kind (Schülerin/Schüler). Wählen wir beispielsweise eines der „Männchen“ im Graph aus, so können wir anhand der aktuellen Datenkarte die Daten zum Kind ablesen (siehe folgendes Beispiel):



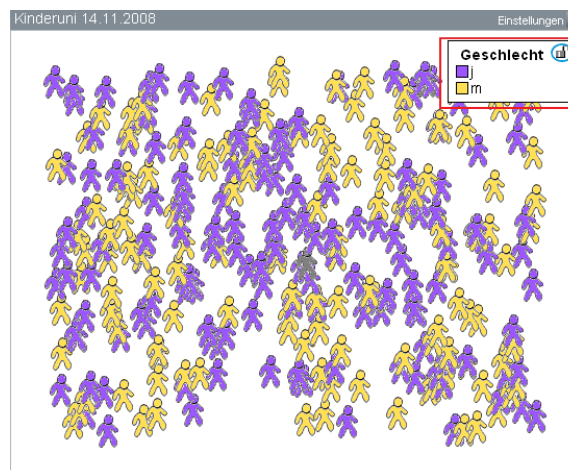
Nun kann man beispielsweise die Daten von „Smily“ ablesen. Klicken wir nun ein Merkmal in der Datenkarte an (exemplarisch das Merkmal Geschlecht), so färben sich die Symbole je nach Geschlecht (männlich: gelb / weiblich: lila) ein. Ebenso kann man mit anderen Merkmalen verfahren.



Will man nun zwischen Jungen und Mädchen unterscheiden können, oder besser gesagt –auf einfachem Wege- herausfinden, welches Geschlecht welcher Farbe zugeordnet ist, so kann sich der Legende-Funktion bedienen. Aber wo finden wir die? Antwort: Durch Erzeugen des Graphs erscheint die Arbeitsleiste:

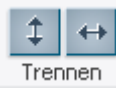


Hier findet man ganz rechts die Funktion „Legende“. Durch Betätigung erscheint im Graph rechts oben (roter Kasten in folgender Abb.) ein Fenster, in welchem wir ablesen können, dass die gelben „Männchen“ weiblich und die „Männchen“ in lila weiblich sind. (Anm.: Was es mit dem Schloss (blauer Kreis in folgender Abb.) auf sich hat, das werden wir später erfahren. Im Moment spielt diese Funktion für uns noch keine große Rolle.)

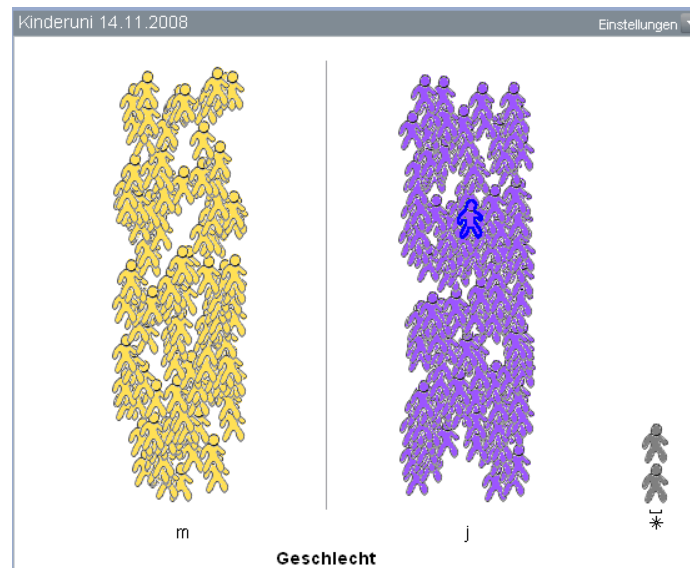


Nun wollen wir Ordnung in die Daten bzw. Datenkarten bzw. „Männchen“ bringen. Dieses machen wir auf einer ersten Stufe, indem wir die Jungen und die Mädchen –ähnlich wie bei dem Arbeiten mit Datenkarten- trennen, um uns einen Überblick über die Verteilung des Merkmals „Geschlecht“ in diesem Datensatz zu verschaffen.

Dieses können wir mit der Funktion „Trennen“ realisieren: entweder durch Betätigung des



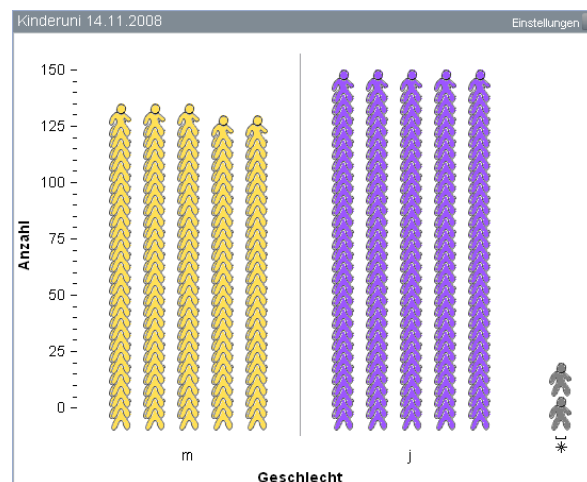
Schalters oder aber durch „Ziehen“ eines „Männchen“ im Graphen.



Nun haben wir zwei Klassen erstellt: In der linken Klasse befinden sich die Mädchen, in der rechten die Jungen (die beiden grauen Symbole am rechten Rand repräsentieren Befragte, die keine Angabe zum Merkmal „Geschlecht“ gemacht haben). Die „Männchen“ sind in beiden Klassen leider noch sehr unsortiert.



Durch die Funktion „Stapeln“ wollen wir die „Männchen“ für eine bessere Übersicht „übereinander legen“.

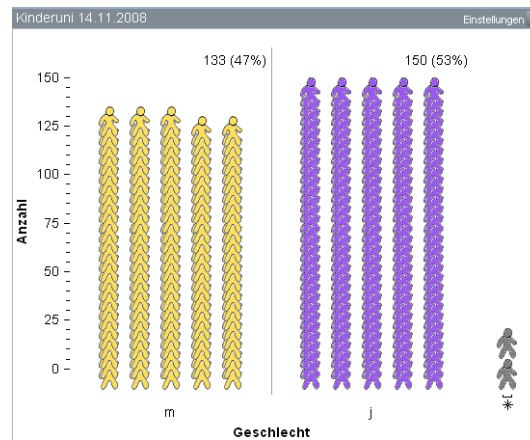


Wir haben nun ein „lupenreines“ Piktogramm erstellt, mit dem wir unsere Frage nach der Geschlechterverteilung in diesem Datensatz nun zumindest qualitativ beantworten können. Wollen wir sie auch quantitativ beantworten, so bedienen wir uns der „Anzahl“-Achse links

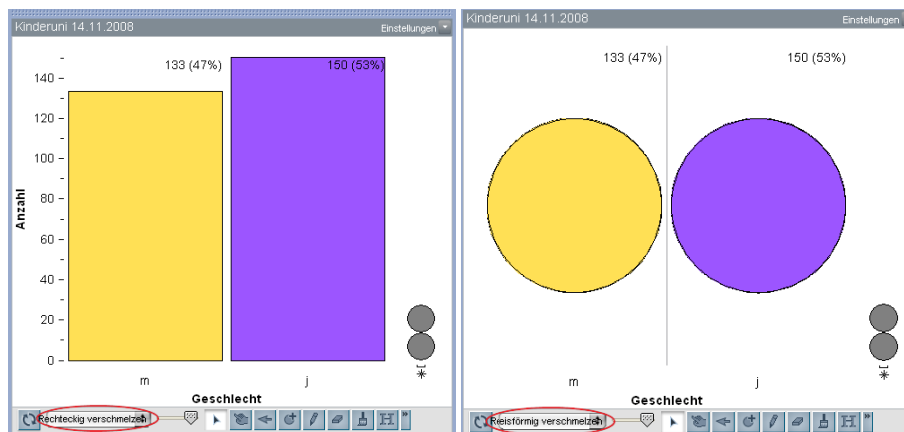
bzw. der Funktion „Anzahl“



. Diese liefert absolute und relative Häufigkeiten für einzelne Klassen.

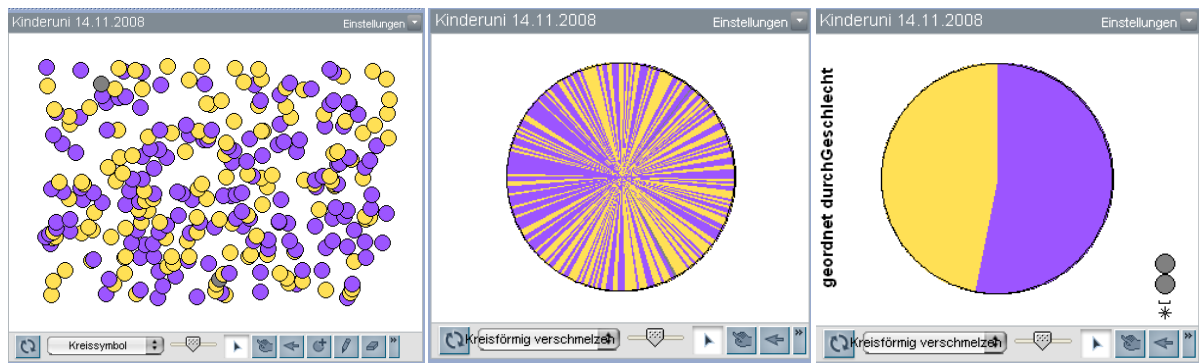


Will man nun vom Piktogramm zu eher konventionellen Graphiken, wie z.B. das Säulendiagramm oder auch Kreisdiagramm überschwenken, so können wir auch dieses realisieren. Zunächst zum Säulendiagramm: Durch Betätigen der Funktion „Rechteckig verschmelzen“ (siehe folgende Abb., rote Ellipse), erhalten wir ein Säulendiagramm. Auf ähnliche Art und Weise können wir auch ein Kreisdiagramm erstellen. Dazu nutzen wir die „verwandte“ Funktion „Kreisförmig verschmelzen“ (siehe folgende Abbildung rechts).

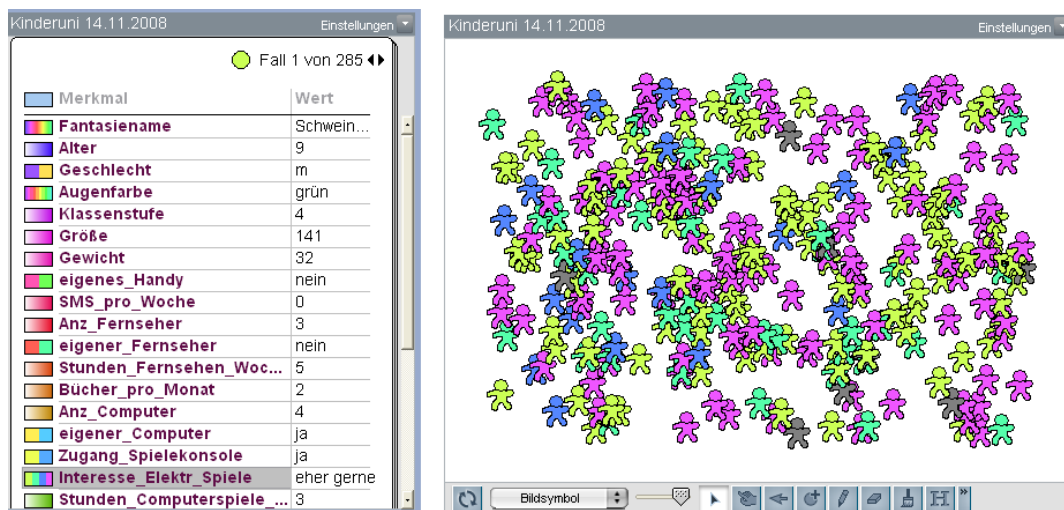


Nun ist dieses Diagramm sicherlich kein Kreisdiagramm im eigentlichen Sinne (denn in diesem Fall hier können wir nur über die eingeblendeten Anzahlen und Anteile etwas über die Verteilung des Merkmals Geschlecht in diesem Datensatz aussagen).

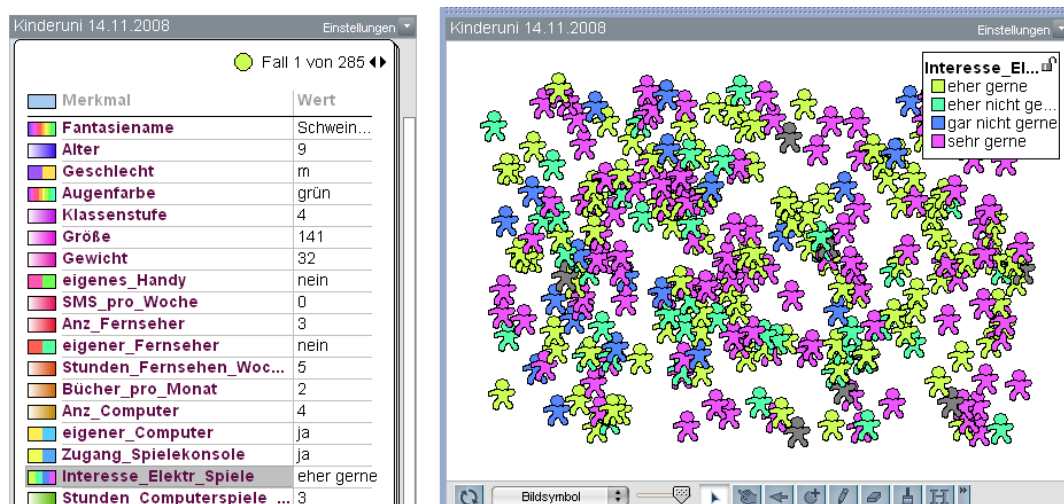
Mit  mischen wir den Graph einmal durch (folgende Abb. links), wählen dann „Kreisförmig verschmelzen“ (folgende Abb., Mitte) und „Ordnen“ dann (folgende Abb. rechts).



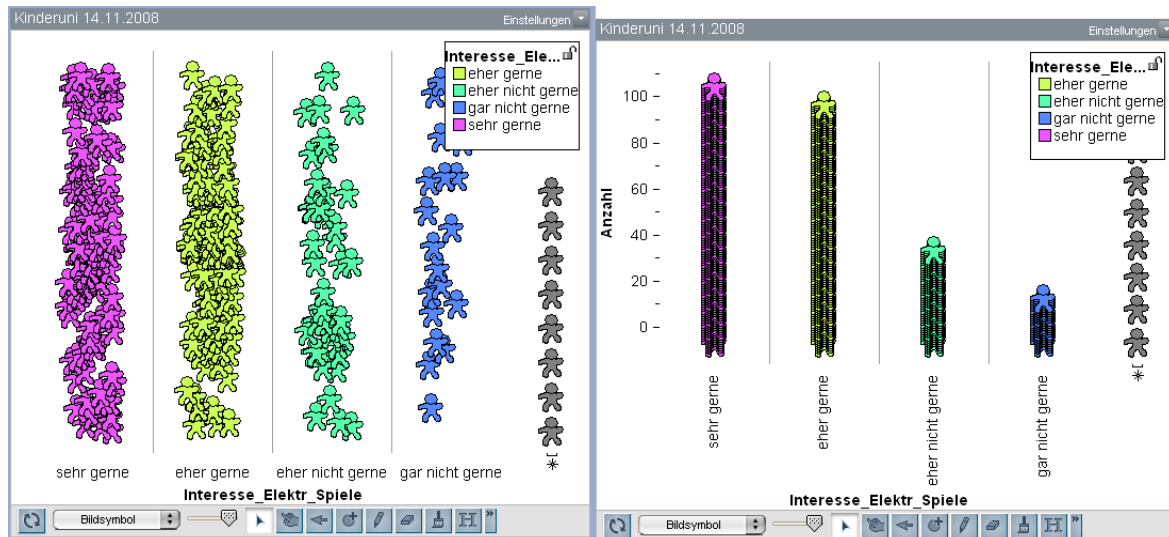
Das waren die ersten Auswertungen in TinkerPlots anhand des Beispiels Merkmal „Geschlecht“. Nehmen wir nun ein qualitatives Merkmal, das mehr als nur zwei Ausprägungen hat, z.B. das Merkmal „Interesse_Elekt_Spiele“ mit den Ausprägungen „sehr gerne“, „eher gerne“, „eher nicht gerne“ und „gar nicht gerne“. Wir wählen das Merkmal in den Datenkarten (links) aus und sehen, dass sich die „Männchen“ gemäß der Ausprägungen einfärben.



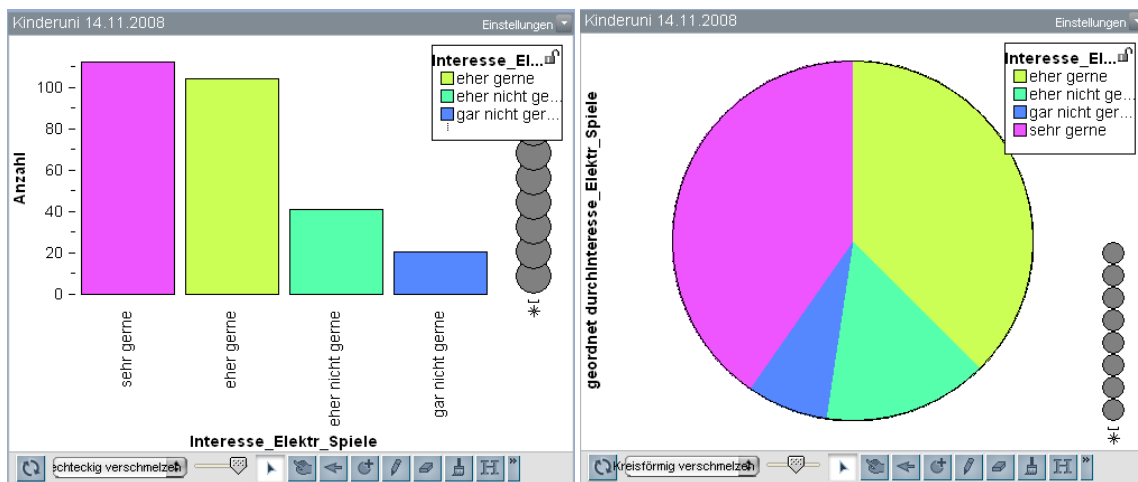
Auch hier hilft im nächsten Schritt wieder das Einblenden unserer Legende:



Wenn wir uns beispielsweise ein Bild darüber machen wollen, wie interessiert die befragten Schülerinnen und Schüler an elektrischen Spielen sind, so müssen wir wieder versuchen Ordnung in die “Männchen” zu bekommen – dieses geschieht wieder mit der „Trennen“ und „Stapeln“-Funktion.

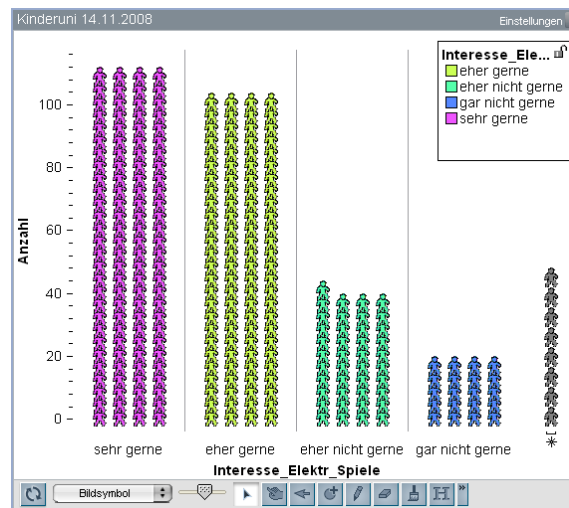


Über die Funktion “Rechteckig verschmelzen” können wir dann wieder ein Säulendiagramm erstellen...oder aber auch ein Kreisdiagramm:

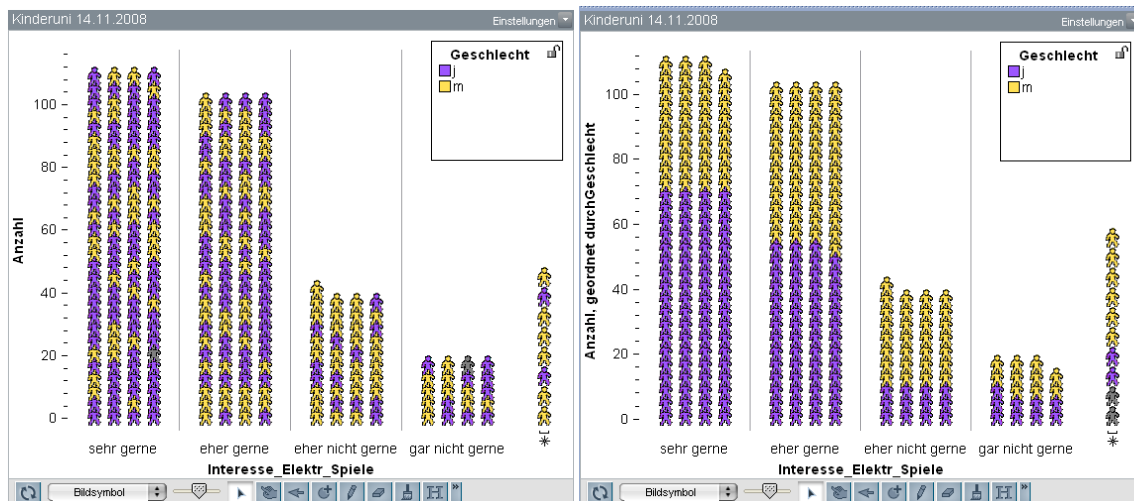


1.4 Erste Verteilungsvergleiche in TinkerPlots

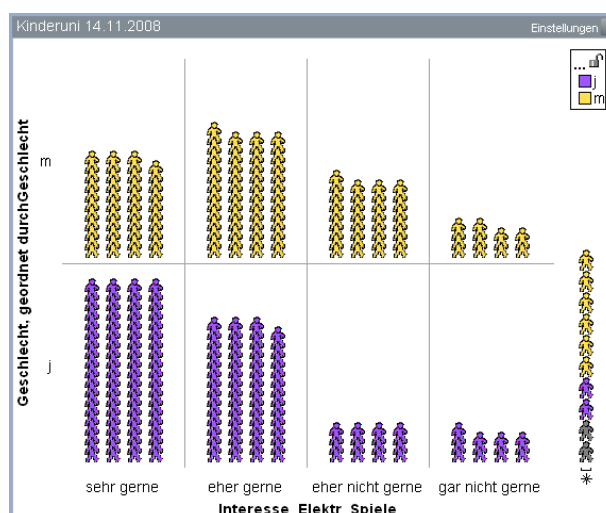
Weiterführend: Es schließen sich einige Fragen an – z.B: Inwiefern unterscheiden sich eigentlich die Jungen und Mädchen in diesem Datensatz bezgl. ihres Interesses an elektrische Spiele? Nehmen wir hier noch einmal unser Piktogramm, dass uns die Verteilung des Merkmals „Interesse_Elekt_Spiele“ für den gesamten Datensatz zeigt:



Wir wählen in den Datenkarten das Merkmal „Geschlecht“ aus: nun werden die Jungen lila und die Mädchen gelb eingefärbt (siehe folgende Abb., links). Auch hier kann “Ordnen” helfen (siehe folgende Abb., rechts).



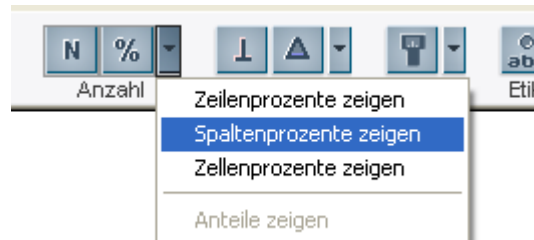
Die Funktion „Trennen“ haben wir bis jetzt nur genutzt, um Klassen auf der „x-Achse“ zu generieren. Wir können aber auch in Richtung der „y-Achse“ trennen, d.h. die Jungen und Mädchen trennen.



Nun haben wir zwei Verteilungen des Merkmals “Interesse_Elekt_Spiele”: zum einen die der Mädchen, zum anderen die der Jungen. Durch Betrachtung der Form lassen sich jetzt schon einige Aussagen treffen, noch genauer können wir sie mit unserer Anzahlfunktion

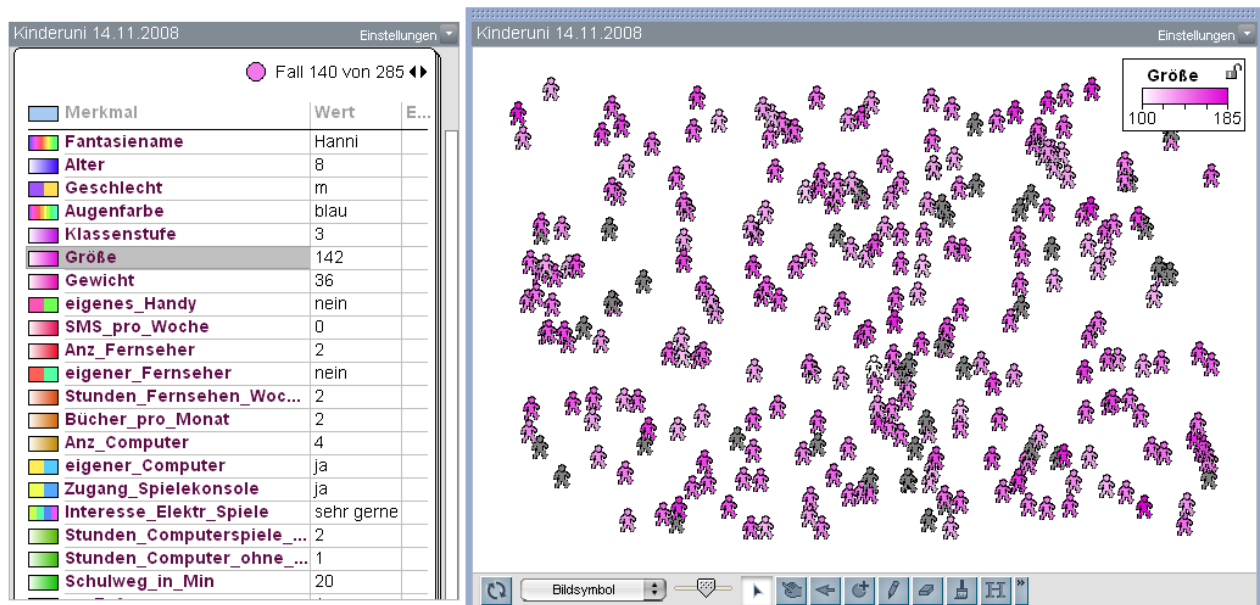


formulieren. Wir können nun mit absoluten Häufigkeiten aber auch mit relativen Häufigkeiten argumentieren. Gerade bei den relativen Häufigkeiten bietet sich weiteres Potential. Im Untermenü findet man hier auch die Möglichkeit sich verschiedene relative

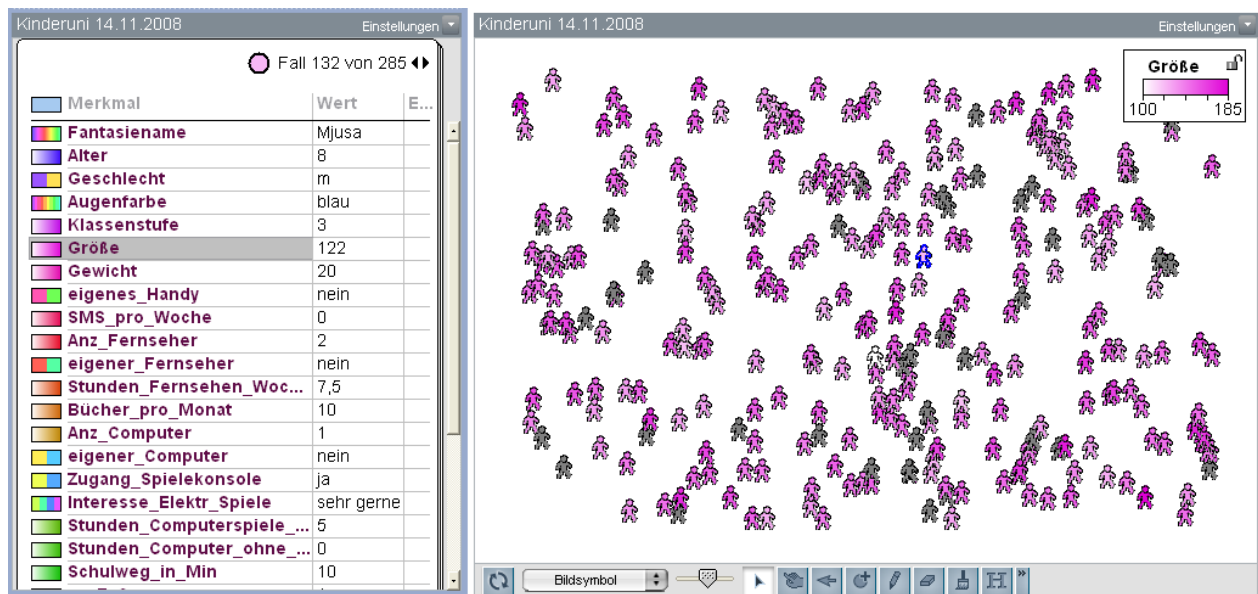


Häufigkeiten anzuzeigen

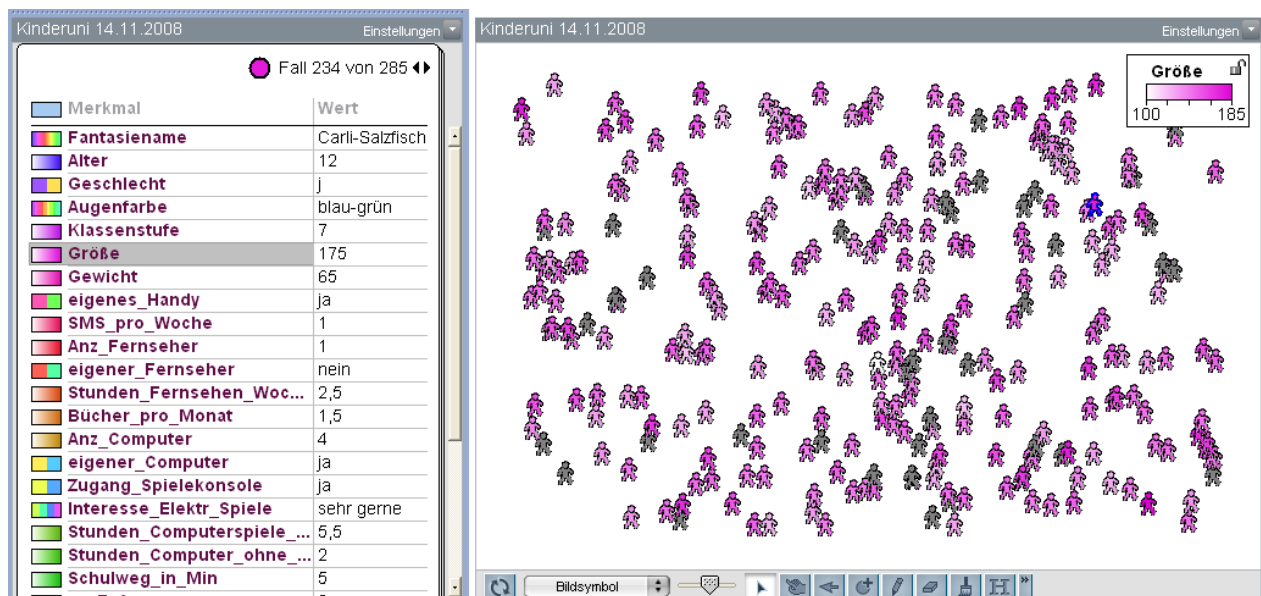
- nämlich: Zeilen-, Spalten- und Zellenprozent. Auf diese wollen wir bei dieser elementaren Einführung allerdings nicht weiter eingehen. - Wir haben bis hierhin nur die Verteilung bzw. Verteilungen von qualitativen bzw. kategorialen Merkmalen (Geschlecht, Interesse_elekt_Spiele) betrachtet: Wie sieht es aber mit quantitativen Merkmalen aus? Wir wählen das Merkmal „Größe“ (Körpergröße) in den Datenkarten aus.



Die “Männchen” sind nun lila eingefärbt, aber in unterschiedlichen Farbintensitäten – hier sehen wir nun wie in TinkerPlots kategoriale und numerische Merkmale anhand der Einfärbung unterschiedlich dargestellt werden. Bei den kategorialen Merkmalen hatten wir noch verschiedenen Ausprägungen verschiedene Farben zugeordnet. Nun -bei numerischen Merkmalen- sind die „Männchen“ in einer Farbe eingefärbt, aber mit unterschiedlichen Intensitäten: desto stärker der Farbton ausgeprägt ist, desto größer ist das jeweilige Kind („Männchen“). Die Verbindung zu den Datenkarten sehen wir sofort, wenn wir zwei Kinder im Graph anklicken.

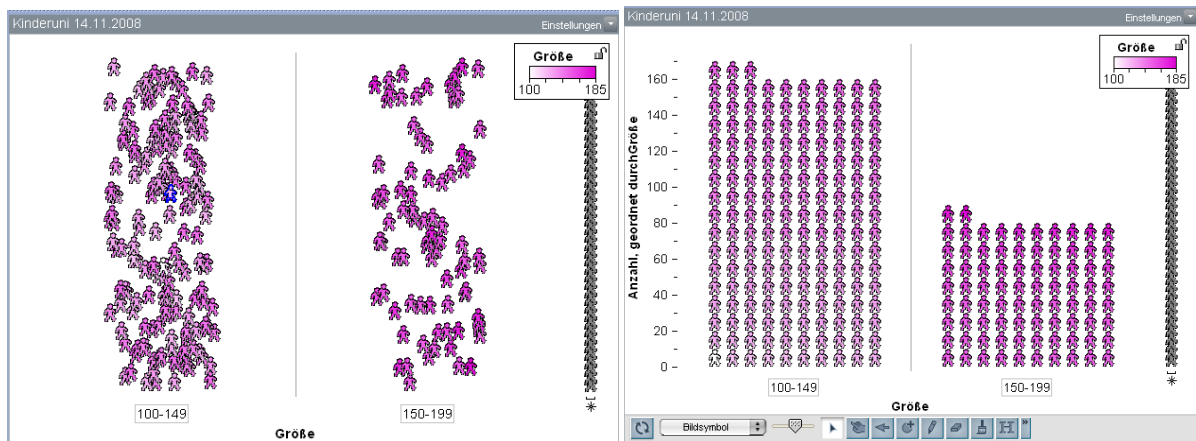


Wir haben “Mjusa” angewählt. Sie ist 122cm groß.

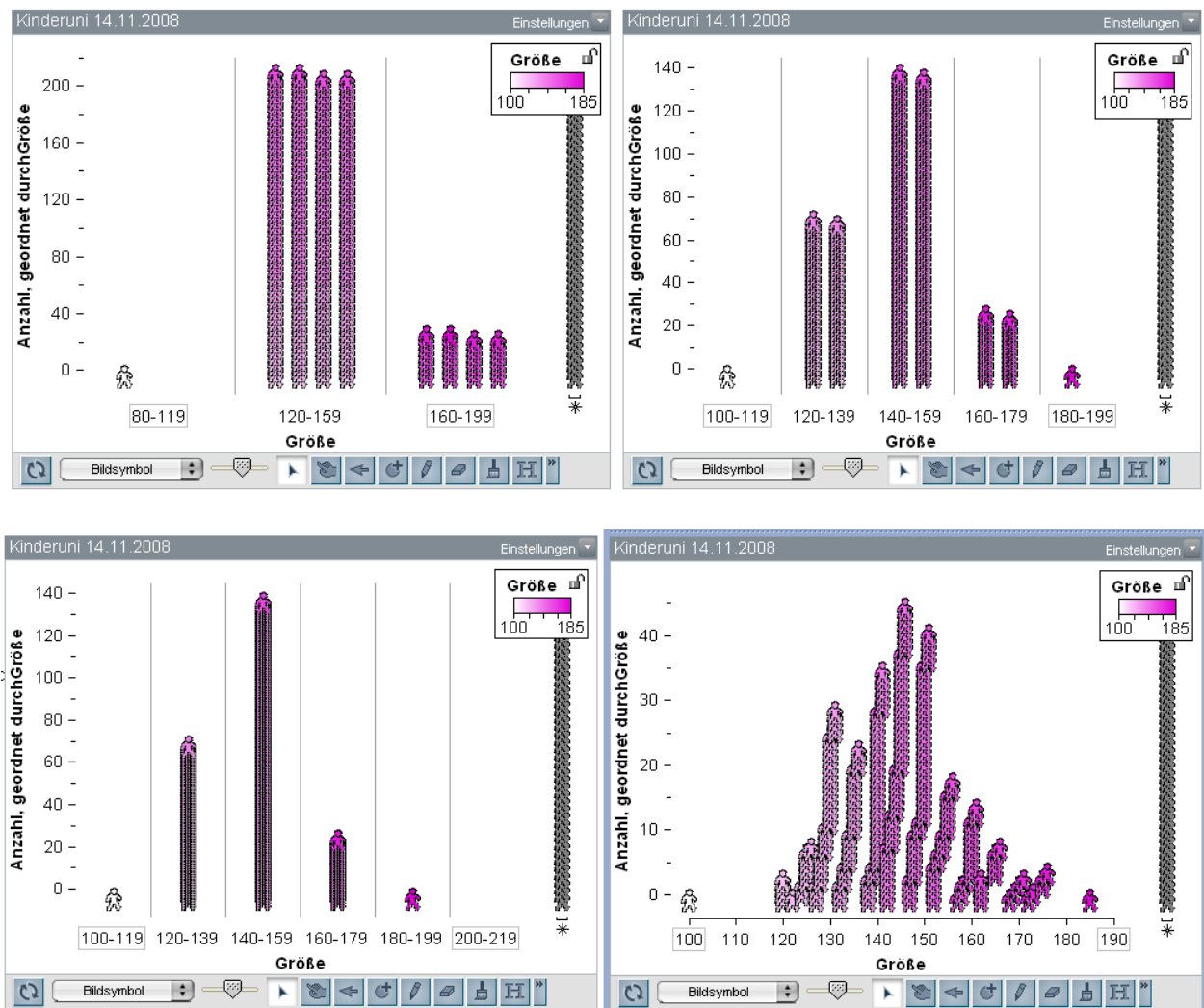


Carli-Salzfisch hingegen ist 175cm groß.

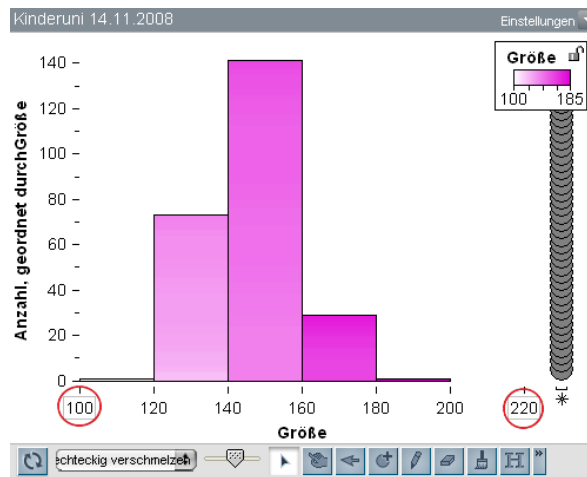
Nun wollen wir eine Übersicht über die Verteilung des Merkmals trennen bekommen – wie gehen wir vor? Wir nutzen wieder unsere Grundoperationen „Stapeln“, „Trennen“ und „Ordnen“, die wir analog auch mit unseren Datenkarten realisieren können. Mithilfe von „Trennen“ teilen wir die Kinder in zwei Klassen hinsichtlich ihrer Körpergröße ein und betrachten die folgende Abbildung (links): In der linken Klasse haben wir nun die Kinder, die zwischen 100cm und 149cm, in der rechten Klasse, die zwischen 150cm und 199cm groß sind. Auch hier kann „Stapeln“ und „Ordnen“ Übersicht verschaffen, wie wir (rechts) sehen können:



Wir abermaliges Trennen können wir die Einteilung der Klassen immer weiter verfeinern, bis wir ein gestapeltes „Punktdiagramm“ haben:

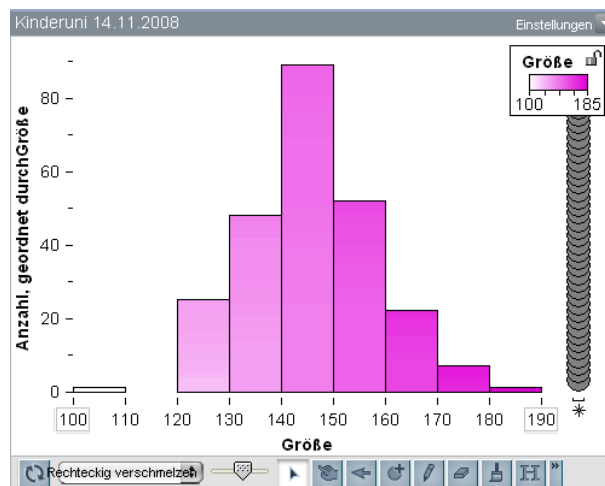


Nehmen wir beispielsweise den vorletzten “Schritt”. Mit dem Befehl „Rechteckig verschmelzen“ kann ein Histogramm mit der Klassenbreite 20 erstellt werden.

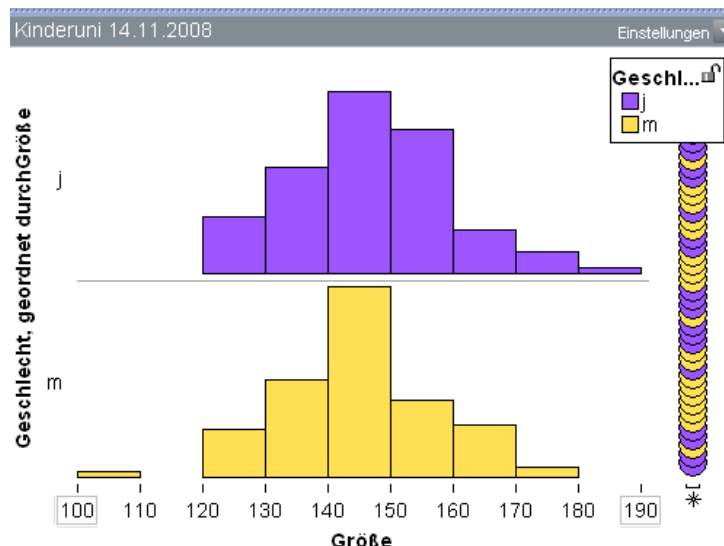


Die Klasseneinteilung (sowie auch der Anfang und das Ende der Achse) kann auch manuell durch Vorgabe geändert werden. Dazu einfach auf einen der Kästen (rote Umrahmung) am linken oder rechten Ende der x-Achse klicken. Im dann erscheinenden Dialogfenster kann dann die Klassenbreite variiert werden:

Man könnte, wie in der folgenden Abbildung z.B. die Klassenbreite 10 wählen.



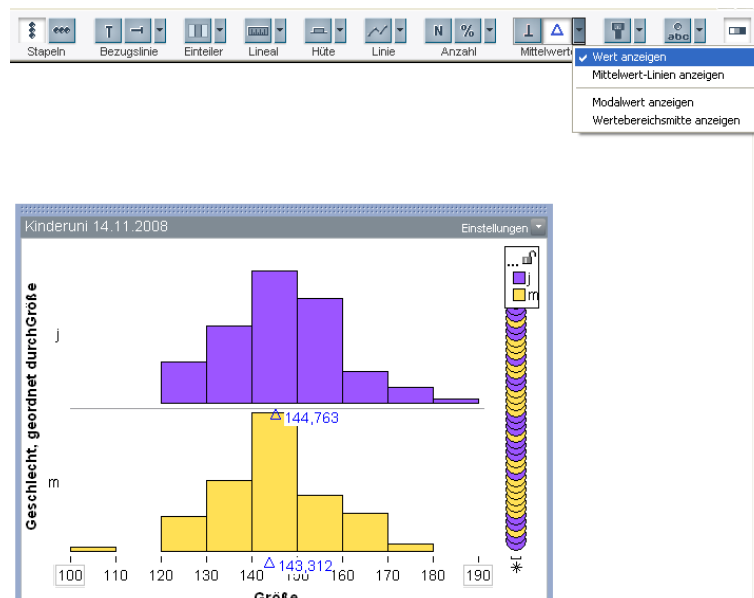
Nun könnte man wieder eine Frage nach Unterschieden aufwerfen, z.B. Unterschiede geschlechtsspezifischer Art, d.h. inwiefern sich die Mädchen und Jungen hinsichtlich des Merkmals Körpergröße unterscheiden. Auch hier können wir genauso vorgehen, wie bei den kategorialen Variablen und zwischen Jungen und Mädchen trennen.



Nun gilt es Unterschiede in den Gruppen herauszuarbeiten, z.B. könnte man mit der Form argumentieren (die in diesem Fall doch recht gleich, symmetrisch, zu sein scheint)³ oder aber auch, wie wir es bereits oben bei der kategorialen Merkmalen gemacht haben mit den „verschiedenen Prozents“ (Zeilen-, Spalten- und Zellenprozente). Man könnte auch Mittelwerte wie den Median oder das arithmetische Mittel verwenden. Diese können unter



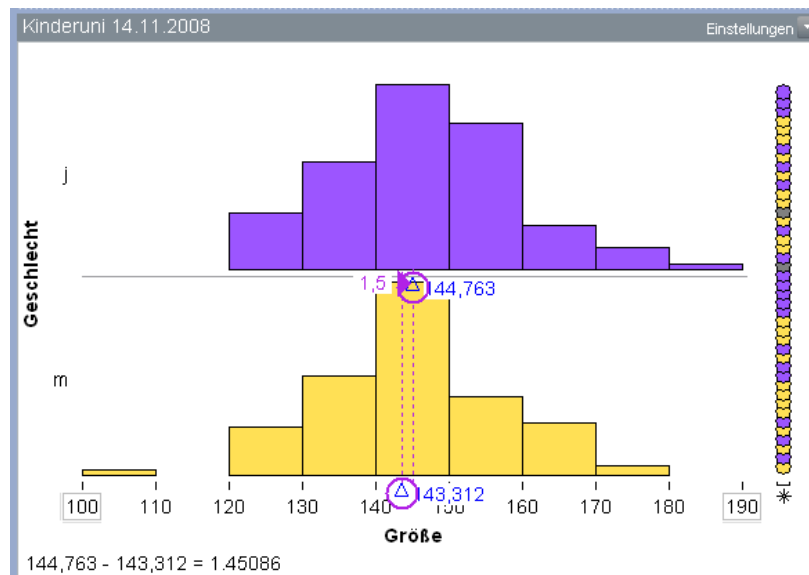
Betätigung von **Mittelwerte** eingeblendet werden. Dabei steht das rote invertierte T für den Median, das blaue Dreieck für das arithmetische Mittel. Wir blenden nun das arithmetische Mittel ein und lassen uns, wie in der folgenden Abbildung zu sehen ist, den Wert dazu anzeigen.



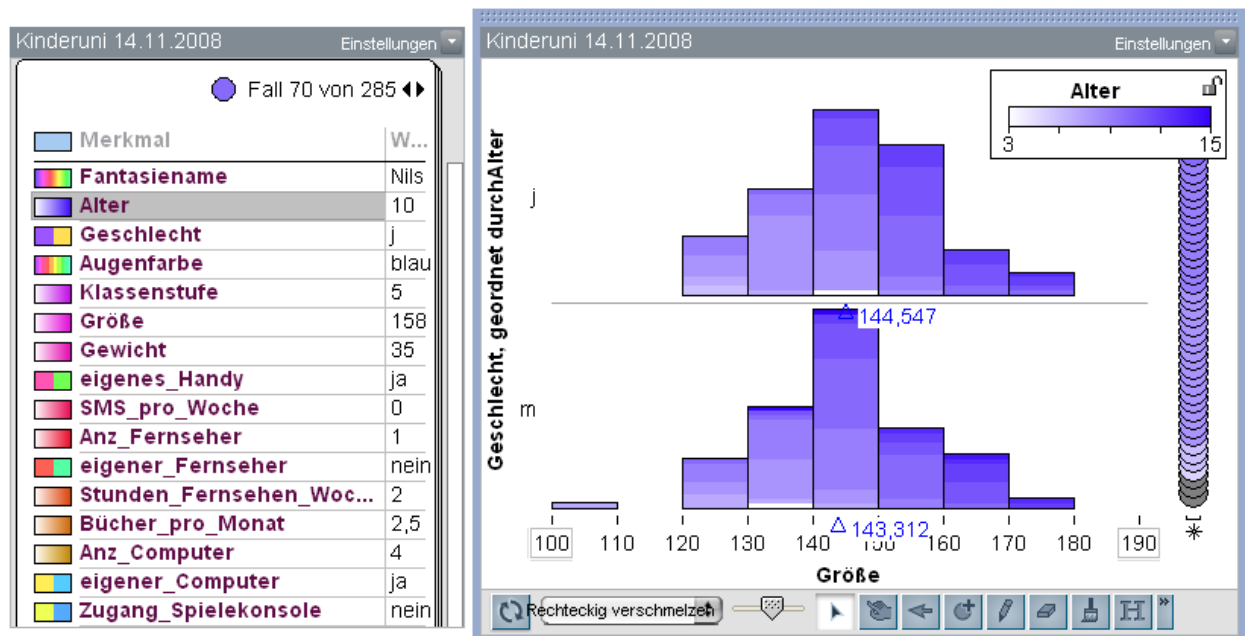
Im Durchschnitt sind die Jungen in diesem Datensatz ca. 145 cm, die Mädchen ca. 143 cm groß. Die Jungen sind hier also tendenziell 2 cm größer als die Mädchen.

³ Man beachte die Gegebenheit der Stichprobe! (Der KinderUni-Datensatz enthält Kinder im Alter zwischen 3 und 15 Jahren!)

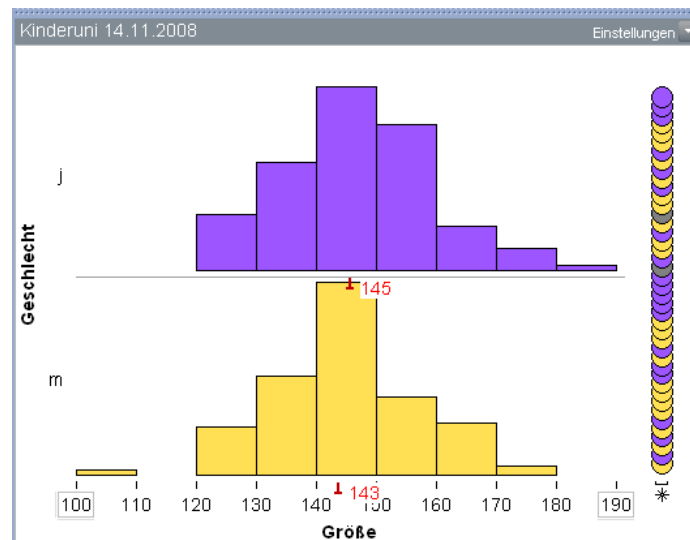
Will man den Abstand genau berechnen, so kann man sich das Lineal zu Hilfe nehmen und den Abstand der beiden arithmetischen Mittelwerte ausrechnen.



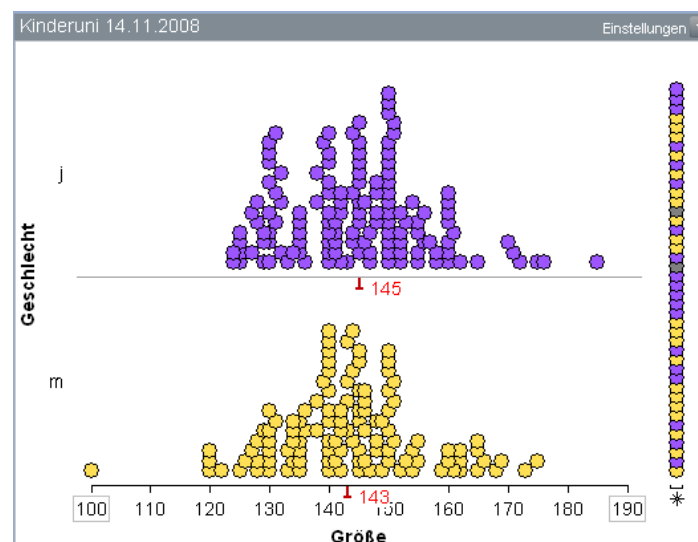
Dieser geringe Unterschied scheint auf dem ersten Blick zu verwundern. Er erklärt sich jedoch, wenn wir uns wie in der Anmerkung oben noch einmal in Erinnerung rufen, welche Kinder befragt worden sind und als weiteren Auswertungsschritt das Merkmal Alter hinzunehmen, d.h. die Säulen nach dem Merkmal Alter einfärben.



Wir können auch den Median berechnen lassen und versuchen diesbezüglich weitere Unterschiede in den beiden Gruppen festzustellen.



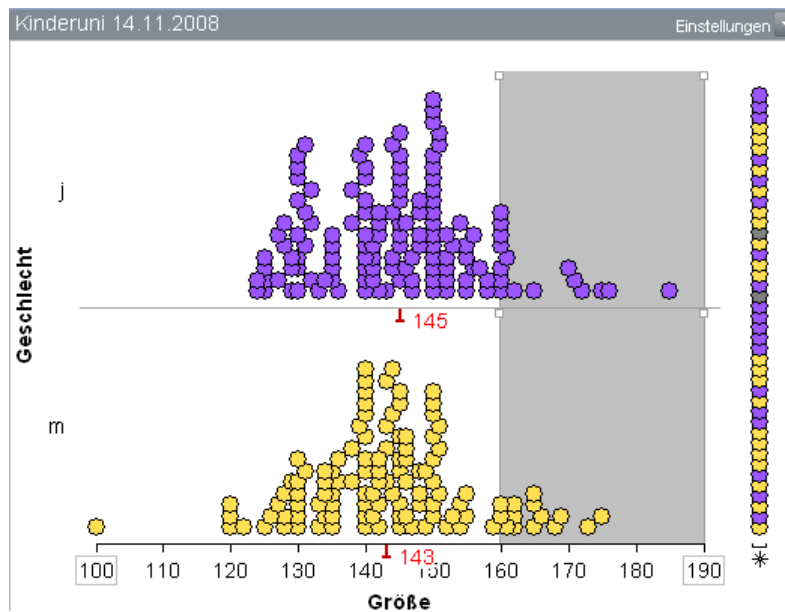
Wir sehen, dass der Median der Verteilung des Merkmals Größe ähnlich wie das arithmetische Mittel bei den Jungen etwas größer ist, als bei den Mädchen. Wir haben in den vorangegangenen Beispielen das Histogramm als Darstellungsform benutzt, ähnliches kann natürlich auch mit einem Punktdiagramm geschehen:



Für uns stellt sich nun die Frage: wie viele Jungen sind größer als 160cm? Wie viele sind das bei den Mädchen? Gibt es mehr größere Jungen als größere Mädchen in unserem Datensatz? Um Fragen solcher Art beantworten zu können, eignen sich insbesondere sogenannte Daten –

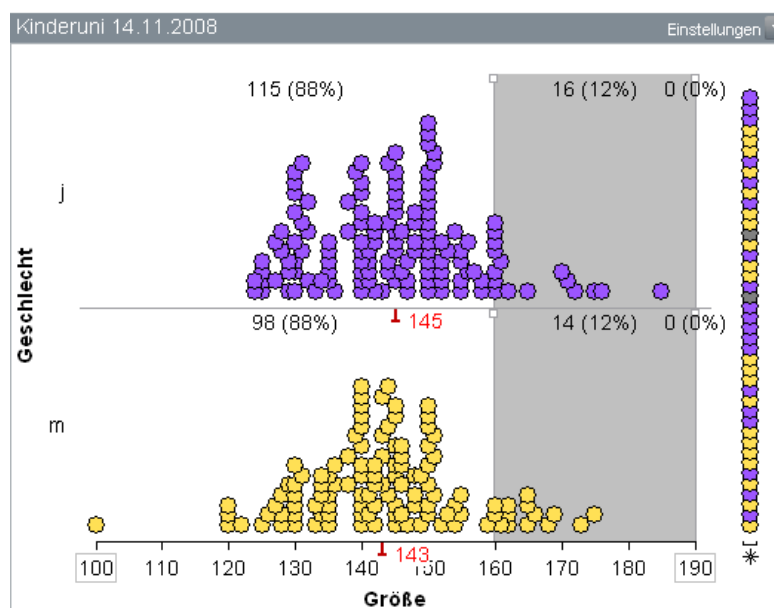


„Einteiler“ Einteiler, die TinkerPlots zur Verfügung stellt. Mit diesem Werkzeug können wir nun z.B. die Anzahl und den Anteil (in der jeweiligen Klasse) der Jungen, die größer oder gleich 160 cm sind, ermitteln:



Einblenden der Anzahlen und Anteile

liefert uns jetzt die Antwort auf unsere Fragen:

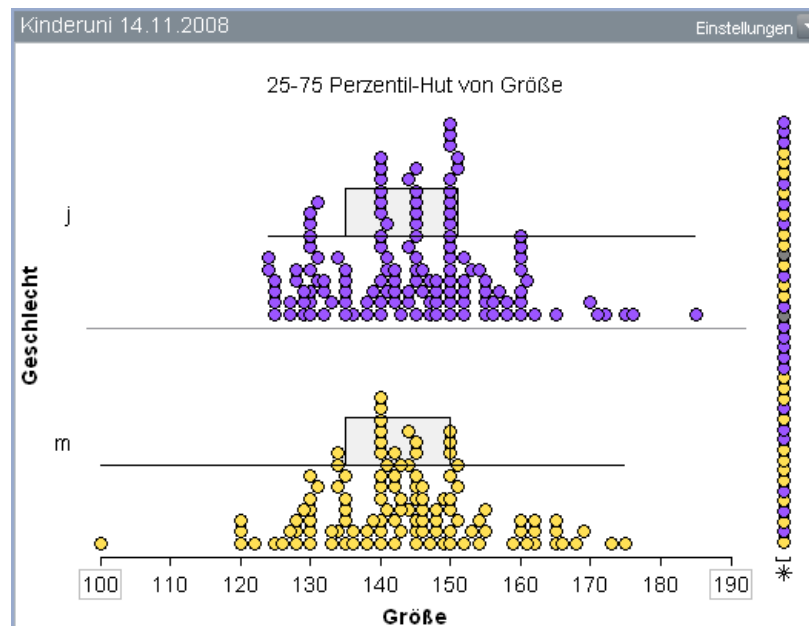


Unsere kleine elementare Einführung in die Datenanalyse mit TinkerPlots soll mit der Vorstellung der Hüte und Boxplots, die ebenfalls grundlegende Instrumente und Darstellungen für den Vergleich zweier Verteilungen sind, abschließen.

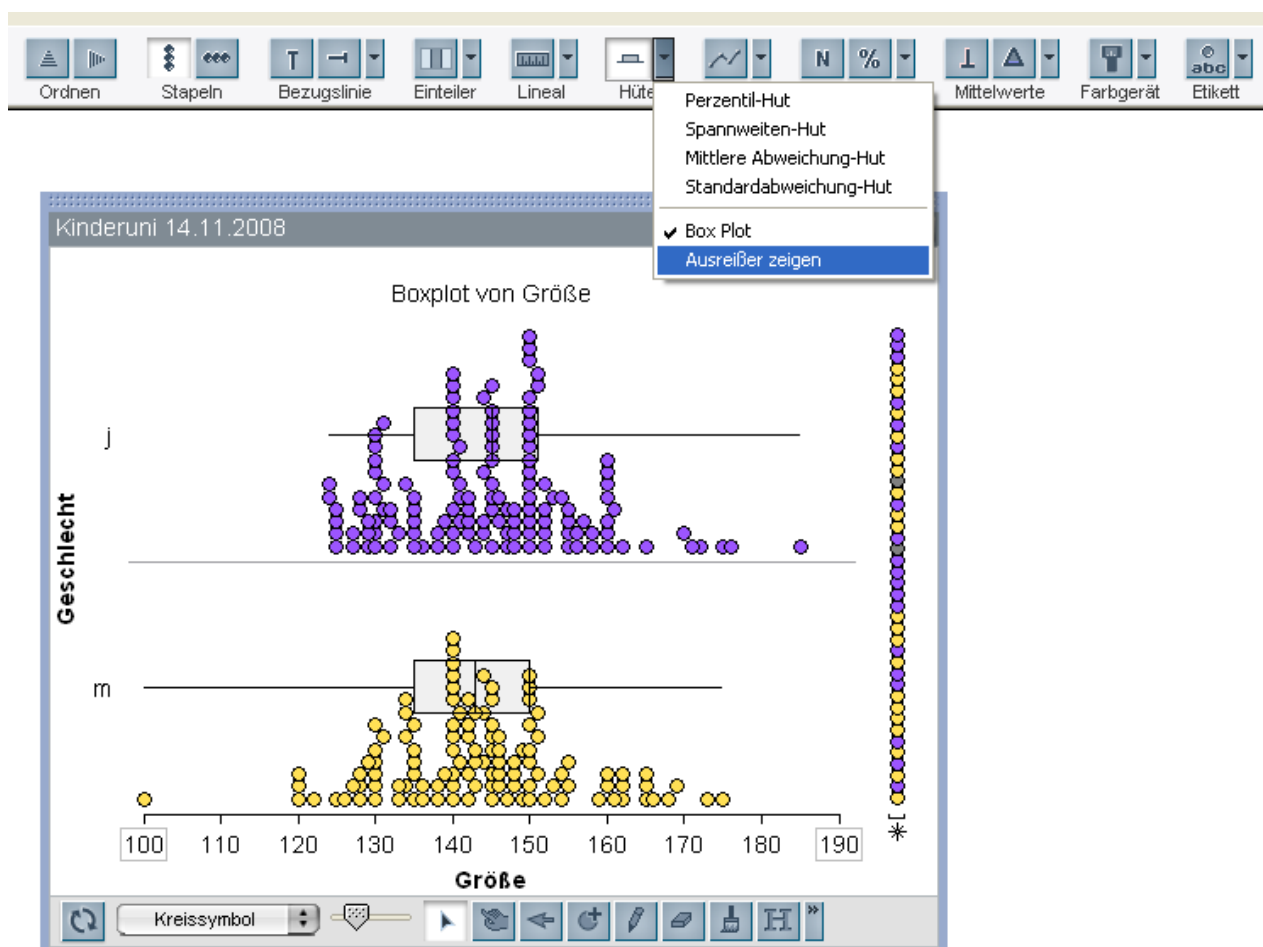
Die Option Hüte



legt beispielsweise einen Hut über die Daten.



Unter der Krone des Hutes liegen die mittleren 50% der Verteilung, an den Rändern bzw. unter der linken Krempe / rechten Krempe liegen 25% der Daten. Dieses schafft die Vorstufe zum Boxplot den wir ebenfalls als Darstellungsform auswählen können:



Auf Wunsch lassen sich auch Ausreißer anzeigen.

erstellen, über den Rechtsklick erhält man die bekannten Operationen „ausschneiden“, „kopieren“ und „einfügen“.

Generell lassen sich in der Datentabelle ähnliche Einstellungen wie in den Datenkarten vornehmen.

2. Stochastische Simulationen mit der Zufallsmaschine in TinkerPlots

2.1 Die Zufallsmaschine in der Übersicht

The screenshot shows the 'Zufallsmaschine' (Random Machine) interface in TinkerPlots. The interface includes a top control bar with 'RUN', pause, and stop buttons, a speed slider set to 'mittel', and an 'Einstellungen' (Settings) dropdown. Below this is a header area with 'Durchgänge 5' (Iterations 5) and 'Merkm1 Merkm2' (Features 1 and 2). The main workspace contains a circular diagram labeled 'Ziehungen 2' (Drawings 2) with two arrows pointing to it. Below the diagram are three circular buttons labeled 'a', 'b', and 'a', with plus, minus, and reset buttons underneath. At the bottom is a toolbar with icons for 'Box', 'Stapel' (Stack), 'Kreisel' (Spinner), 'Balken' (Bar), 'Kurve' (Curve), and 'Zähler' (Counter). The 'Zähler' icon shows a table with values 1, 2, and 3.

Die Anzahl der Ziehungen legt die Anzahl der Merkmale fest

Die Steuerung der Zufallsmaschine befindet sich in der obersten Leiste: Starten (RUN), pausieren, stoppen; Einstellen der Geschwindigkeit

So viele **Durchgänge** (Wiederholungen) gibt es pro Simulationsdurchlauf

So oft wird pro Durchgang gezogen:

Zum **Wechseln** ein Bauteil in die Mitte ziehen

Die Inhalte des Bauteils können mit Hilfe dieser Buttons verändert werden: Elemente hinzufügen, entfernen, etc.

Über das **Pfeilmenü** kann z. B. „ohne Zurücklegen“ eingestellt werden.

2.2 Die einzelnen Bauteile der Zufallsmaschine



Die **Box** kann mit maximal 10.000 beliebig beschrifteten Kugeln gefüllt werden. Mit dem  und  Button werden Kugeln hinzugefügt, bzw. gelöscht. Die Kugeln können beliebig beschriftet werden. Bei einer großen Anzahl an Kugeln (oder Zeichen pro Kugel) ist dies jedoch recht mühsam. Liegen die Elemente in einer Liste (z. B. in Excel) oder als Textdatei in einzelnen Zeilen vor, so lassen sie sich einfach durch Kopieren und Einfügen in den Mixer legen. Zahl- oder Buchstabenbereiche lassen sich komfortabel über das Eingabefenster  eingeben. Beim Starten (Klick auf Run) wird zufällig eine Kugel ausgewählt, ihr Wert entsprechend der Ziehung unter den Merkmalsnamen geschrieben und sie je nach Einstellung zurückgelegt. Es kann mit oder ohne Zurücklegen gezogen werden, standardmäßig wird mit Zurücklegen gezogen.



Beim Bauteil **Stapel** können maximal 10.000 Stapel gebildet werden. Dabei kann die Höhe jedes einzelnen Stapels beliebig variiert werden, die entsprechende absolute Anzahl kann über das Pfeilmenü eingeblendet werden. Mit der Beschriftung verhält es sich wie bei der Box. Beim Starten wird zufällig ein Element eines Stapels ausgewählt und dessen Wert unter das entsprechende Merkmal geschrieben. Es kann mit oder ohne Zurücklegen gezogen werden. Standardmäßig wird mit Zurücklegen gezogen. Benannte Stapel dürfen auch 0 Elemente enthalten.



Am **Kreisel** können maximal 10.000 Segmente erstellt werden, wobei die Größe jedes einzelnen individuell einstellbar ist (in der Summe müssen sich 100 Prozent ergeben). Die entsprechende Prozentzahl oder der entsprechende Anteil kann über das Pfeilmenü eingeblendet werden. Mit der Beschriftung verhält es sich wie bei den beiden vorherigen Bauteilen. Hier kann nur mit Zurücklegen gezogen werden.



Im Bauteil **Balken** gibt es für jedes Element einen Balken, der in der Höhe angepasst werden kann. Die jeweilige Prozentzahl oder der Anteil der einzelnen Balken kann über das Pfeilmenü eingeblendet werden, in der Summe müssen sich auch hier 100 %, bzw. 1 ergeben. Auch hier sind maximal 10.000 Balken möglich. Für die Beschriftung gilt das Gleiche wie bisher. Läuft die Zufallsmaschine, so wird zufällig innerhalb eines Balkens ein Ort angewählt und der entsprechende Balkenwert gezogen. Hier kann nur mit Zurücklegen gezogen werden.



Das Bauteil **Kurve** weicht nun etwas von den vorherigen ab. Hiermit können nur numerische Werte erzeugt werden. Die Grenzen der Kurve ergeben ein Intervall, auf dem die Fläche einer Kurve liegt. Die einzige Beschränkung dabei ist, dass es sich um reelle Zahlen handeln muss. Wird die linke Zahl größer als die rechte gewählt, so werden die Grenzen automatisch vertauscht. Der Verlauf der Kurve ist beliebig veränderbar, allerdings ist es oft mühsam, den gewünschten Kurvenverlauf einzuzeichnen. Eine Funktionseingabe ist nicht möglich. Läuft die Zufallsmaschine wird ein Wert innerhalb der Fläche unter der Kurve zufällig ausgewählt

und die so ermittelte reelle Zahl dem entsprechenden Merkmal zugeschrieben. Die Höhe der Kurve gibt dabei die Wahrscheinlichkeit an, mit der die reelle Zahl gezogen wird. Auch hier kann nur mit Zurücklegen gezogen werden.



Das letzte Element, der **Zähler**, nimmt eine Sonderrolle ein. Hierbei handelt es sich nicht um ein Zufallselement, sondern um einen linearen Stückzähler. Nach jedem Zug wird das Rad genau ein Feld weitergedreht, bis es wieder von vorne anfängt. Auch hier können maximal 10.000 Einträge vorgenommen werden, bei beliebiger Beschriftung. Hier ist nur Ziehen mit Zurücklegen möglich.

3. Ausgewählte Aufgaben zur...

3.1 Datenanalyse mit TinkerPlots

- Einführung in die Datenanalyse mit TinkerPlots anhand von Datenkarten
- Exemplarische Vorstellung der Software am Datensatz „KinderUni“:
 - Einfache statistische Auswertungen und erste Darstellungen mit TinkerPlots – Umgang mit Häufigkeitsverteilungen
 - Verteilungsvergleiche mit TinkerPlots
mögliche Beispielfragen:
 - Haben Schüler der Sek I einen zeitlich längeren Schulweg als Grundschüler?
 - Wie unterscheiden sich Handy- und Nicht-Handy-Besitzer hinsichtlich ihres Gewichts?

3.2 Simulation mit TinkerPlots

Demonstrationsaufgaben zur Simulation:

1. „Doppelwurf“

Jemand bietet dir ein Würfelspiel an. Dazu sollen zwei Würfel gleichzeitig geworfen und die Augensumme gezählt werden. Du darfst dir vorher aussuchen, ob Du mit den Augensummen 5, 6, 7, 8 (Ereignis A) oder mit allen übrigen Augensummen (Ereignis B) gewinnen möchtest. Begründe, ob du eine der beiden Gewinnmöglichkeiten bevorzugen würdest.

Optional 2. „Chevalier de Méré“

Obwohl es nach meinen Überlegungen beim Werfen von drei Würfeln gleich viele Möglichkeiten für die Augensumme 11 und die Augensumme 12 gibt, ist es meiner Erfahrung nach besser, auf die Augensumme 11 zu setzen. Warum?

3. „Pflanze raten“

Welche Pflanze versteckt sich hier? Mit wie vielen Durchgängen errätst Du die Pflanze?

Aufgabe „Doppelwurf“

Jemand bietet dir ein Würfelspiel an. Dazu sollen zwei Würfel gleichzeitig geworfen und die Augensumme gezählt werden. Du darfst dir vorher aussuchen, ob Du mit den Augensummen 5, 6, 7, 8 (Ereignis A) oder mit allen übrigen Augensummen (Ereignis B) gewinnen möchtest. Begründe, ob du eine der beiden Gewinnmöglichkeiten bevorzugen würdest.⁴

3.2.1 Modellierung des „doppelten Würfelwurfs“ mit der Zufallsmaschine



Erläuterung: In einer Box liegen 6 Kugeln, beschriftet von 1 bis 6, aus dieser wird zweimal mit Zurücklegen gezogen (Anzahl Ziehungen). Dieses zweistufige Zufallsexperiment wird 1000 Mal wiederholt (Anzahl Durchgänge).

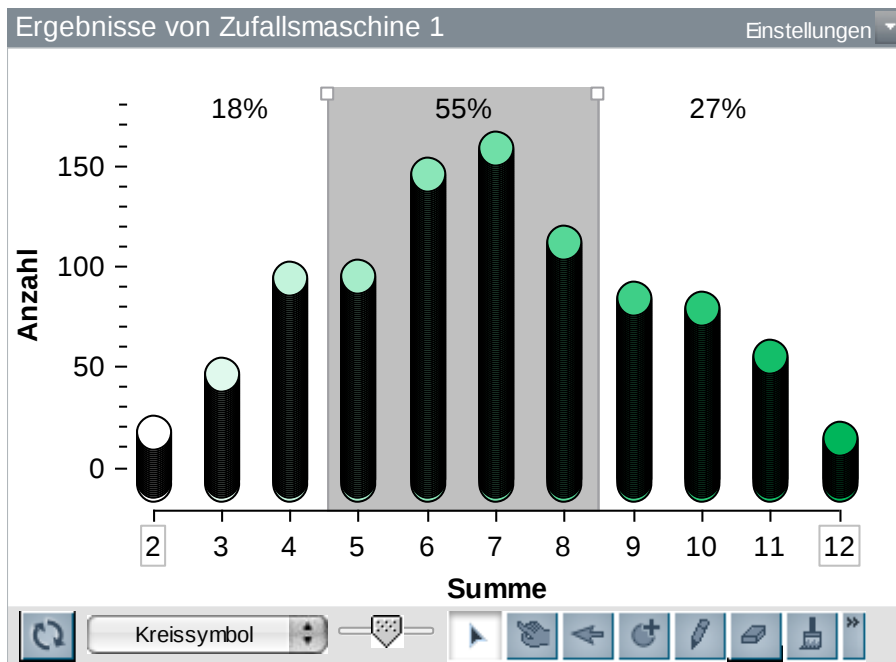
Definieren der Zufallsgröße „Augensumme“

Ergebnisse von Zufallsmaschine 1					Einstellungen
	Gesamt	Merkm1	Merkm2	Summe	
Formel	SummeElemente (Gesamt)				
					
996	3;1	3	1	4	
997	1;5	1	5	6	
998	2;4	2	4	6	
999	6;1	6	1	7	
1000	6;5	6	5	11	

Erläuterung: Dargestellt ist in der ersten Spalte „Gesamt“ die Kombination beider Ziehungen, anschließend gibt es für jede Ziehung ein eigenes Merkmal („Merkm1“ und „Merkm2“). In der letzten Spalte „Summe“ wird die Summe der Elemente in „Gesamt“ gebildet. Diese Spalte kann durch einfaches Auswählen in einem Menü erzeugt werden. Hier ist die Summe die interessierende Größe, gebildet als die Summe der zwei Würfelwürfe.

Darstellen der Zufallsgröße und Interpretation

⁴ Aus: Müller, J.H. (2005): Die Wahrscheinlichkeit von Augensummen – Stochastische Vorstellungen und stochastische Modellbildung. In: Praxis der Mathematik in der Schule, Heft 3 (47). S. 17 – 22.



Dargestellt ist die Verteilung der Augensumme von zwei Würfelwürfen bei 1000 Durchgängen. Mit Hilfe des Einteilers ist das Ereignis A (Summe 5, 6, 7 und 8) gekennzeichnet und die Prozentzahl eingeblendet.

Bei 1000 doppelten Würfelwürfen tritt das Ereignis A „Die Augensumme 5, 6, 7 oder 8 kommt vor“ in ca. 55 % der Fälle auf, das heißt, die gesuchte Wahrscheinlichkeit wird mit 95 % Sicherheit im Intervall $[0,52; 0,58]$ liegen. Das Ereignis A hat also eine höhere Wahrscheinlichkeit ($>50\%$) als das Ereignis B und sollte deshalb bevorzugt werden.

3.2.2 Die Tabelle und ihre Formeln am Beispiel des „doppelten Würfelwurfs“

Beispiel: Doppelter Würfelwurf

	Gesamt	Merkm1	Merkm2	<neu>
1	4;5	4	5	
2	6;4	6	4	
3	3;6	3	6	
4	6;6	6	6	
5	1;2	1	2	

Die Ergebnisse der Zufallsmaschine werden in einer Tabelle protokolliert. In der Tabelle gibt es immer das Merkmal „Gesamt“, das alle Ergebnisse eines Durchgangs zusammenfasst und die einzelnen Merkmale als Spalten entsprechend der Anzahl der Ziehungen. Hinweis: Wurde in der Zufallsmaschine nur eine Ziehung eingestellt, so gibt es insgesamt nur eine Spalte „Merkm1“.

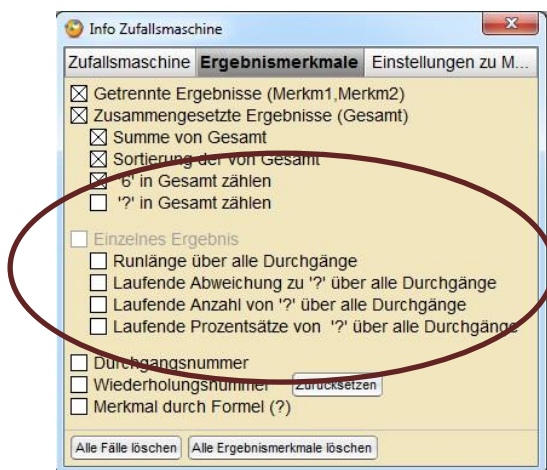
Jeder Durchgang der Zufallsmaschine bekommt eine Zeile. Es gibt also so viele Zeilen in der Tabelle wie Durchgänge in der Zufallsmaschine.

Merke: Zeilen sind Durchgänge; Spalten sind Ziehungen

Im Beispiel: Gesamt bezeichnet beide Würfelergebnisse eines Durchgangs, Merkm1 kann als der erste Wurf interpretiert werden und Merkm2 als der zweite Wurf. Beim ersten doppelten Würfelwurf (Zeile 1) wurde erst eine 4, dann eine 5 geworfen.

Auswertungen in der Tabelle:

Der Tabelle können zusätzliche Spalten hinzugefügt werden, normalerweise werden damit Auswertungen realisiert. Spalten können von Hand angefügt werden oder über ein Menü. Mit einem Klick auf den Pfeil rechts oben neben *Einstellungen* kann das folgende Fenster geöffnet werden, mit dem einfache Auswertungen als weitere Spalten automatisch erstellt werden (Eintrag *Ergebnismerkmale*).



Ergebnisse von Zufallsmaschine 1						
	Gesamt	Merkm1	Merkm2	Summe	Kombinationen	Anzahl_6
1	4;5	4	5	9	4;5	0
2	6;4	6	4	10	4;6	1
3	3;6	3	6	9	3;6	1
4	6;6	6	6	12	6;6	2
5	1;2	1	2	3	1;2	0

Erklärung dieser „vordefinierten Formeln“:

Summe von Gesamt: Bildet die Summe der Einträge der Zelle Gesamt. Ist nur sinnvoll bei numerischen Einträgen

Sortierung von Gesamt: Sortiert die Einträge der Zelle Gesamt aufsteigend. Funktioniert bei numerischen und bei kategorialen Einträgen

„?“ in Gesamt zählen: Zählt wie häufig ein bestimmtes Element in Gesamt vorkommt. Als Beispiel wurde hier die 6 gewählt. Kann auf numerische und kategoriale Merkmale angewendet werden

Eigene Formeln verwenden:

Alternativ können Formeln mit Hilfe eines Formeleditors eingegeben werden. Dazu muss zunächst die Formelzeile in der Tabelle eingeblendet werden. Dies funktioniert mit einem Rechtsklick auf die Tabelle und den Eintrag *Formeln zeigen*.

The screenshot shows a window titled "Ergebnisse von Zufallsmaschine 1". It contains a table with columns: Gesamt, Merkmal1, Merkmal2, and a button labeled "<neu>". The table has 5 rows of data. A right-click context menu is open over the table, with the option "Formeln zeigen" highlighted.

	Gesamt	Merkmal1	Merkmal2	<neu>
1	4;5	4	5	
2	6;4	6	4	
3	3;6	3	6	
4	6;6	6	6	
5	1;2	1	2	

Context menu options:

- Neue Fälle...
- Neues Merkmal...
- Einfügen
- Fälle alle auswählen
- Einheiten zeigen
- Formeln zeigen**
- Farben verbergen
- Fallsymbole zeigen
- Verborgene Merkmale anzeigen
- Automatisches Anpassen der Spaltenbreiten
- Zufall erneuern (Ctrl+Y)
- In eigenem Fenster zeigen (Alt+V)
- Tabelle vervielfältigen (Ctrl+D)
- Tabelle löschen (Delete)

Soll beispielsweise die Differenz (als Betrag) der beiden Würfelergebnisse berechnet werden, so kann zunächst ein neues Merkmal mit „Differenz“ benannt werden. Ein Doppelklick auf die grau unterlegte Formelzeile öffnet den Formeleditor.

The screenshot shows the same table as before, but with a new column "Differenz" added. A double-click arrow points to the "Formel" row in the "Differenz" column. The "Formel für Differenz" dialog box is open, showing a formula editor with a list of available features and functions.

	Gesamt	Merkmal1	Merkmal2	Differenz
Formel				
1	4;5	4	5	
2	6;4	6	4	
3	3;6	3	6	
4	6;6	6	6	
5	1;2	1	2	

Formel für Differenz dialog box:

Differenz =

groß

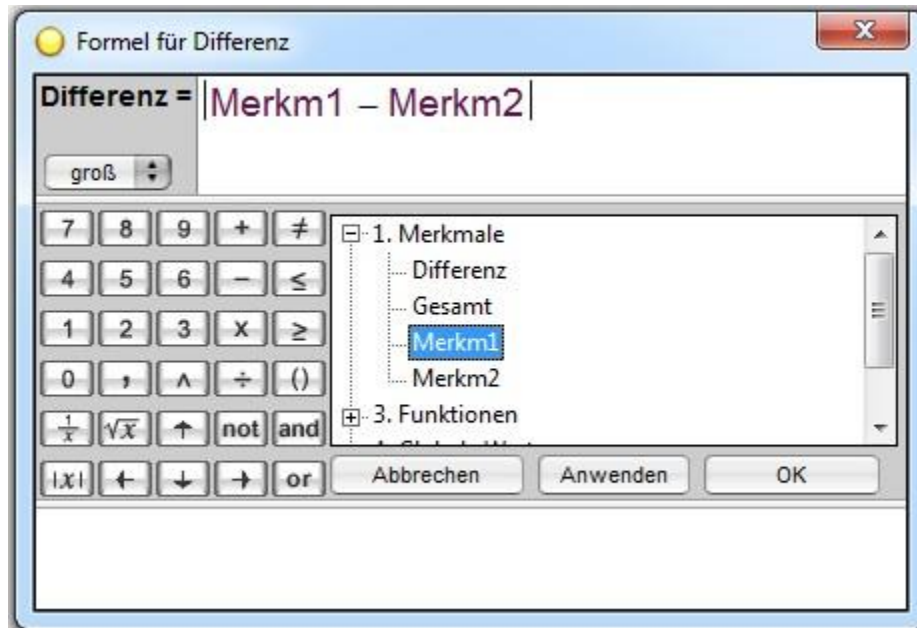
Available features:

- 1. Merkmale
 - Differenz
 - Gesamt
 - Merkmal1
 - Merkmal2
- 3. Funktionen

Buttons: Abbrechen, Anwenden, OK

Text: Merkmalsnamen können in Ausdrücken verwendet werden; sie beziehen sich auf die zugehörigen Merkmale in der Kollektion.

Im Formeleditor können nun Formeln eingegeben werden. Dabei gibt es eine bestimmte Syntax, die es zu beachten gilt. Es finden sich viele Befehle im Formeleditor: Im oberen Feld wird die Formel/der Befehl eingegeben, im mittleren rechten Feld findet sich die Auflistung sämtlicher verfügbarer Befehle und im unteren Feld befindet sich eine Erklärung (oft mit Beispielen) zu dem gewählten Befehl. Im Fall der Differenz kann die Formel von Hand eingegeben werden. Hilfreich dabei ist das mittlere linke Auswahlfeld, durch Drücken der STRG-Taste ändern sich die Elemente, bspw. Wird aus = ein $\neq$. Auch die Striche für Beträge findet man dort.



Ergebnisse von Zufallsmaschine 1					Einstellungen
	Gesamt	Merkm1	Merkm2	Differenz	
Formel				Merkm1 - Merkm2	
1	4;5	4	5	1	
2	6;4	6	4	2	
3	3;6	3	6	3	
4	6;6	6	6	0	
5	1;2	1	2	1	

Wie bei den anderen Merkmalen zuvor findet sich in der neuen Spalte „Differenz“ das Ergebnis passend zu der jeweiligen Zeile.

Alle Merkmale können in Graphen dargestellt, weiter untersucht und ausgewertet werden.

