

Tabellen und Grafiken mit Excel und Calc

Inhalt

| | |
|--|----|
| 1. Tabellen und Grafiken in Excel | 2 |
| 1.1 Daten sortieren und filtern..... | 2 |
| 1.2 Eine Spalte mit Zufallszahlen hinzufügen | 5 |
| 1.2 Tabellen..... | 7 |
| 1.3 Grafiken..... | 9 |
| 2. Tabellen und Grafiken mit Calc | 11 |
| 2.1 Daten sortieren und filtern..... | 11 |
| 2.2 Eine Spalte mit Zufallszahlen hinzufügen | 14 |
| 2.2 Tabellen..... | 15 |
| 2.3 Grafiken..... | 18 |

In diesem Tutorial geht es darum, wie man Tabellenkalkulationsprogramme insbesondere für korpuslinguistische Zwecke nutzt. Vieles lässt sich natürlich auch auf andere Bereiche übertragen. Insbesondere werden wir uns mit der Frage befassen, wie man Daten in Tabellen zusammenfassen und grafisch darstellen kann. Wie in den vorherigen Tutorials werden wir sowohl das kommerzielle Microsoft Excel als auch sein kostenloses Pendant LibreOffice Calc benutzen.

Wie man CSV-Spreadsheets in Excel und Calc einliest, haben wir bereits im Tutorial 1 („Grundlegendes“) gelernt und werden daher an dieser Stelle nicht näher darauf eingehen. Stattdessen gehen wir *in medias res* und werten eine Beispieltabelle mit fiktiven Daten aus, die sich im Ordner „Excel_Einstieg“ befindet und den Namen *anreden.csv* trägt. Sie enthält in der Spalte „Anrede“ Lexeme, mit denen zwei GesprächspartnerInnen A und B einander anreden und die in der Spalte „Konnotation“ als *positiv*, *negativ* oder *neutral* annotiert sind.

Das Wichtigste zuerst: Autokorrektur abschalten!

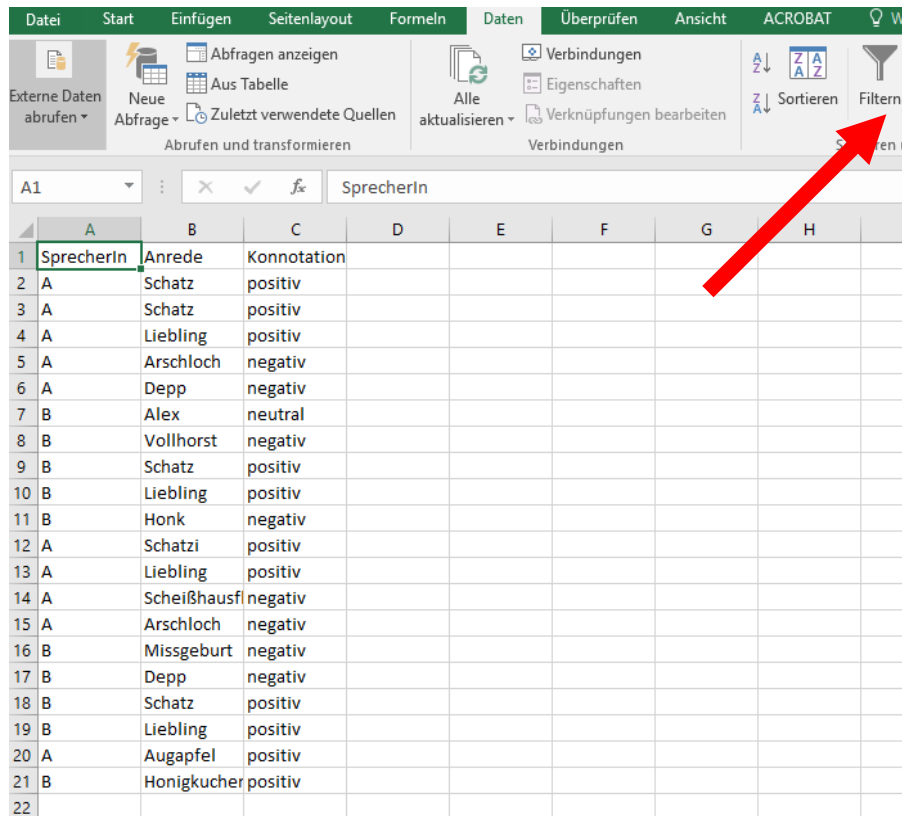
Gerade bei der Arbeit mit historischen Daten ist es extrem wichtig, in Text- und Datenverarbeitungsprogrammen, insbesondere jedoch in den Office-Programmen, alle Autokorrekturoptionen abzuschalten, die historische Schreibkonventionen als Fehler erkennen und automatisch „korrigieren“ könnten. Das betrifft so ziemlich alle Autokorrektur-Optionen, die in Word und Excel standardmäßig aktiviert sind. Sie finden sie bei Windows unter Datei > Optionen > Dokumentprüfung > Autokorrektur-Optionen und bei Mac unter Word > Einstellungen > Autokorrektur. Ich empfehle, alles zu deaktivieren außer dem Ersetzen von Anführungszeichen durch typographische Anführungszeichen. Auch das Ersetzen von Zeichenkombinationen durch Smileys würde ich vorsichtshalber abschalten. (Ich persönlich mache das immer als allererstes. Ich neige nämlich dazu, Doppelpunkte nach schließender Klammer zu setzen, und wenn ich das nicht deaktiviere, sehen meine Texte immer so traurig aus.)

1. Tabellen und Grafiken in Excel

Hinweis: Die folgenden Ausführungen beziehen sich auf Excel 2016. Das meiste sollte auf andere Versionen wie die Vorgängerversion Excel 2013 übertragbar sein. Unterschiede zwischen PC und Mac habe ich versucht zu berücksichtigen.

1.1 Daten sortieren und filtern

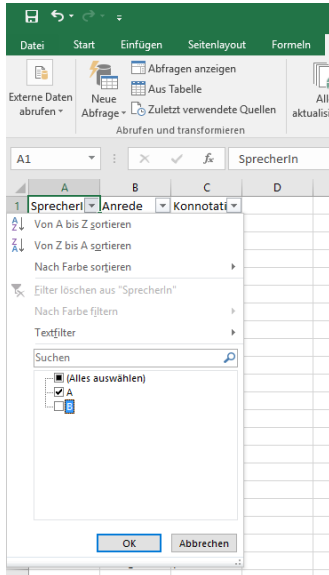
In sehr vielen Fällen kann es nützlich sein, Daten zu sortieren (z.B. alphabetisch) oder zu filtern. Um von diesen Möglichkeiten Gebrauch machen zu können, können wir den Befehl „Filter“ im Reiter „Daten“ verwenden:



Voraussetzung dafür, dass das funktioniert, ist, dass in der ersten Zeile die Spaltennamen stehen (was hier der Fall ist). Bevor wir auf „Filtern“ klicken, stellen wir sicher, dass irgendeine Zelle in der ersten Zeile ausgewählt ist (wie es oben im Screenshot der Fall ist). Sobald wir auf „Filtern“ geklickt haben, erscheint neben jedem Spaltennamen ein kleines Pfeilsymbol:

| | A | B | C |
|---|------------|-----------|-------------|
| 1 | SprecherIn | Anrede | Konnotation |
| 2 | A | Schatz | positiv |
| 3 | A | Schatz | positiv |
| 4 | A | Liebling | positiv |
| 5 | A | Arschloch | negativ |
| 6 | A | Depp | negativ |

Sehen wir uns nun zunächst an, wie man Daten **sortiert**. Angenommen, wir wollen nur Daten von Sprecherin A und nur positiv konnotierte oder neutrale Anreden sehen. Dann können wir zunächst auf das Pfeilchen neben *SprecherIn* klicken und das Häkchen bei „B“ entfernen. Nun werden uns nur noch Daten von Sprecherin A angezeigt. Wichtig: Die Daten von Sprecherin B sind nach wie vor in der Datei vorhanden, sie sind nur ausgeblendet!



Wenn wir nun auf das Häkchen neben „Konnotation“ klicken, sehen wir, dass uns nur zwei Kategorien angezeigt werden: „positiv“ und „negativ“. Die Kategorie „neutral“ wird nicht angezeigt, weil sie nur in den ausgeblendeten Daten vorkommt! Wenn wir nun nur die positiv konnotierten Daten sehen wollen, genügt es, das Häkchen bei „negativ“ zu entfernen.

Trial and error

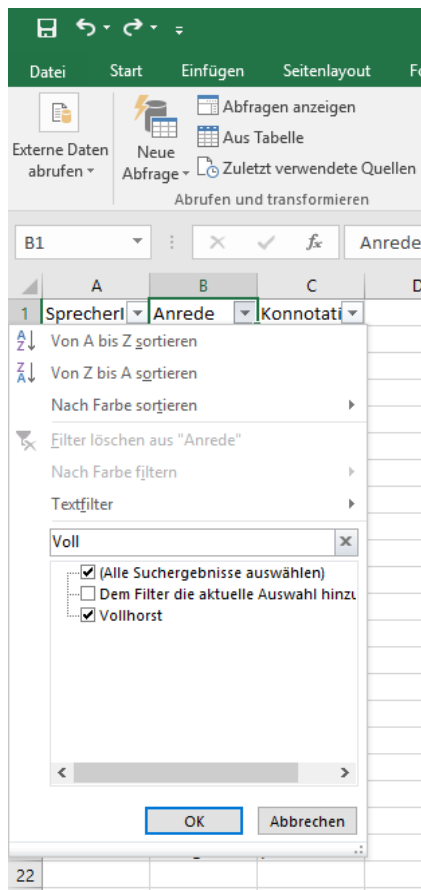
Was passiert, wenn wir nun „B“ wieder einblenden?

Wenn wir nun „B“ wieder einblenden, sehen wir nach wie vor nur „positive“ Belege - die Kategorie „neutral“ müssen wir manuell wieder einblenden, wenn wir alle positiven und neutralen Daten von A und B haben wollen!

Stolperfalle!

Wenn Sie mit mehreren Filtern arbeiten und einen der Filter verändern, kontrollieren Sie bei Bedarf die anderen Filter, um zu verhindern, dass Sie (z.B. bei der Annotation) Daten übersehen, die für Sie relevant sind.

Gerade bei größeren Datensätzen kann es jedoch sinnvoll sein, nicht jede einzelne Variablenausprägung durch Setzen oder Entfernen von Häkchen auszuwählen, sondern stattdessen mit dem Textfilter zu arbeiten, mit dem ich beispielsweise alle Anreden suchen kann, die mit *Voll-* beginnen:



Wie Sie bereits in den obigen Screenshots gesehen haben, haben Sie darüber hinaus die Möglichkeit, die Daten aufsteigend bzw. absteigend zu sortieren. Das kann insbesondere dann sinnvoll sein, wenn Sie Daten haben, in denen einige Types mehrfach vorkommen und bei denen Sie sicher sein können, dass alle Tokens desselben Types dieselbe Annotation bekommen: Dann können Sie einfach das erste Token annotieren und die Annotation dann in alle anderen Zeilen copy&pasten (oder sogar schon vorher alle Zeilen, in die die Annotation kommen soll, markieren, und dann mit Strg+Eingabetaste die eingegebene Annotation auf alle markierten Zeilen übertragen).

Ein Beispiel: Wir wollen unsere positiven und negativen Anreden, bei denen es sich zum großen Teil um Metaphern (oder z.T. auch Metonymien sowie Grenzgängern zwischen beiden) handelt, auf ihren Quellbereich hin annotieren. Dann ist es sinnvoll, die Daten alphabetisch zu sortieren, sodass wir die Tokens, die zum gleichen Type gehören, zusammen annotieren können:

| | Sprecherl | Anrede | Konnotation | Quellbereich |
|----|-----------|--------------|-------------|--------------|
| 2 | B | Alex | neutral | NA |
| 3 | A | Arschloch | negativ | Körperteil |
| 4 | A | Arschloch | negativ | |
| 5 | A | Augapfel | positiv | |
| 6 | A | Depp | negativ | |
| 7 | B | Depp | negativ | |
| 8 | B | Honigkuchen | positiv | |
| 9 | B | Honk | negativ | |
| 10 | A | Liebling | positiv | |
| 11 | B | Liebling | positiv | |
| 12 | A | Liebling | positiv | |
| 13 | B | Liebling | positiv | |
| 14 | B | Missgeburt | negativ | |
| 15 | A | Schatz | positiv | |
| 16 | A | Schatz | positiv | |
| 17 | B | Schatz | positiv | |
| 18 | B | Schatz | positiv | |
| 19 | A | Schatzi | positiv | |
| 20 | A | Scheißhausfl | negativ | |
| 21 | B | Vollhorst | negativ | |
| 22 | | | | |
| 23 | | | | |

Im obigen Screenshot sind die Zellen D3 bis D5 markiert. Es gibt zwei Möglichkeiten, das zu erreichen:

- Zelle D3 markieren und bei gedrückter Umschalt-Taste¹ die Pfeil-nach-unten-Taste drücken.
- Zelle D3 markieren und das kleine Quadrat, das Sie unten rechts an der Markierung sehen, nach unten ziehen. Gerade wenn Sie sehr viele Zellen markieren möchten, kann das manchmal der effizientere Weg sein,

Die Eingabe erfolgt nach wie vor in der ersten Zeile. Wenn Sie nach der Eingabe nur die Eingabetaste drücken, wird Ihre Annotation nicht in die übrigen markierten Spalten übertragen - das geschieht nur, wenn Sie Strg+Eingabetaste drücken.

Alternativ können wir auch *Körperteil* in Zelle D3 schreiben, dann Zelle D3 markieren und die Annotation durch Ziehen am kleinen Quadrat unten rechts an der Markierung auf die beiden folgenden Zeilen übertragen.

Ganz nebenbei sehen Sie im obigen Screenshot übrigens auch eine Konvention, wie man mit Daten umgehen kann, auf die die Annotationskriterien nicht zutreffen und die man deshalb bei der Annotation auslassen möchte: Da *Alex* ein Eigenname ist und daher keinen Quellbereich hat, habe ich ihn mit NA (= Not Available) annotiert.

1.2 Eine Spalte mit Zufallszahlen hinzufügen

Manchmal wollen wir aber die Daten gerade *nicht* nach Kriterien sortieren, die sich aus den Daten selbst ergeben, sondern vielmehr zufällig. Ein Beispiel: Wenn wir diachrone Daten annotieren und die Hypothese haben, dass diachron eine bestimmte Veränderung auftreten wird, dann wäre es sehr problematisch, die Daten nach Jahren sortiert zu annotieren, denn es könnte sein, dass wir bei der Annotation unsere Hypothese zu sehr im Hinterkopf haben und in Zweifelsfällen daher - bewusst oder unbewusst - eher dazu neigen, so zu entscheiden, dass das Ergebnis unserer Hypothese entspricht. In solchen Fällen kann es sinnvoll sein, eine Spalte mit Zufallszahlen hinzuzufügen. Die Formel

=ZUFALLSZAHN()

generiert eine Zufallszahl zwischen 0 und 1. Wenn wir diese Formel in der ersten Zeile anwenden und dann in alle anderen Zeilen übertragen (wiederum indem wir am kleinen

¹ Die Umschalt- oder Shift-Taste ist die Taste ganz links, die Sie sonst benutzen, um Großbuchstaben zu schreiben.

Allerdings: Die Zufallszahlen werden jedes Mal neu generiert. Wenn Sie das Dokument schließen und wieder öffnen, dann haben Sie neue Zufallszahlen und damit auch eine neue Reihenfolge. In der Regel wollen Sie das nicht. Aber auch hier hilft ein Trick. Wenn Sie in eine beliebige Zelle der „Zufallszahl“-Spalte klicken, sehen Sie, dass oben in der Formelspalte noch unsere Formel steht:

| | A | B | C | D | E |
|----|-----------|--------------|-------------|-------------|---|
| 1 | Sprecher1 | Anrede | Konnotation | Zufallszahl | |
| 2 | A | Schatz | positiv | 0.03779296 | |
| 3 | B | Missgeburt | negativ | 0.91402513 | |
| 4 | B | Liebling | positiv | 0.95082988 | |
| 5 | B | Liebling | positiv | 0.22079348 | |
| 6 | A | Liebling | positiv | 0.79959855 | |
| 7 | A | Liebling | positiv | 0.99463391 | |
| 8 | A | Augapfel | positiv | 0.65276906 | |
| 9 | B | Honk | negativ | 0.96448128 | |
| 10 | B | Honigkucher | positiv | 0.91083994 | |
| 11 | A | Schatzi | positiv | 0.73176698 | |
| 12 | B | Alex | neutral | 0.01467371 | |
| 13 | B | Vollhorst | negativ | 0.88494317 | |
| 14 | A | Depp | negativ | 0.75724827 | |
| 15 | B | Depp | negativ | 0.76326371 | |
| 16 | A | Arschloch | negativ | 0.44404618 | |
| 17 | A | Scheißhaufen | negativ | 0.04120787 | |
| 18 | B | Schatz | positiv | 0.90973786 | |
| 19 | B | Schatz | positiv | 0.07942544 | |
| 20 | A | Schatz | positiv | 0.69226583 | |
| 21 | A | Arschloch | negativ | 0.64734578 | |

Das bedeutet, dass hinter den Werten, die wir in der Spalte sehen, immer noch der Zufallsgenerator steht. Jedoch können wir die Kopierfunktion nutzen, um in eine andere (oder auch in dieselbe) Spalte nicht die Formel einzufügen, die wir kopiert haben, sondern die Werte, die sie hervorbringt. Dafür markieren wir zunächst die Zellen, die wir kopieren wollen, und kopieren sie mit Strg + C oder Rechtsklick > Kopieren in die Zwischenablage. Statt sie jedoch einfach mit der Tastenkombination Strg+V einzufügen, wählen wir

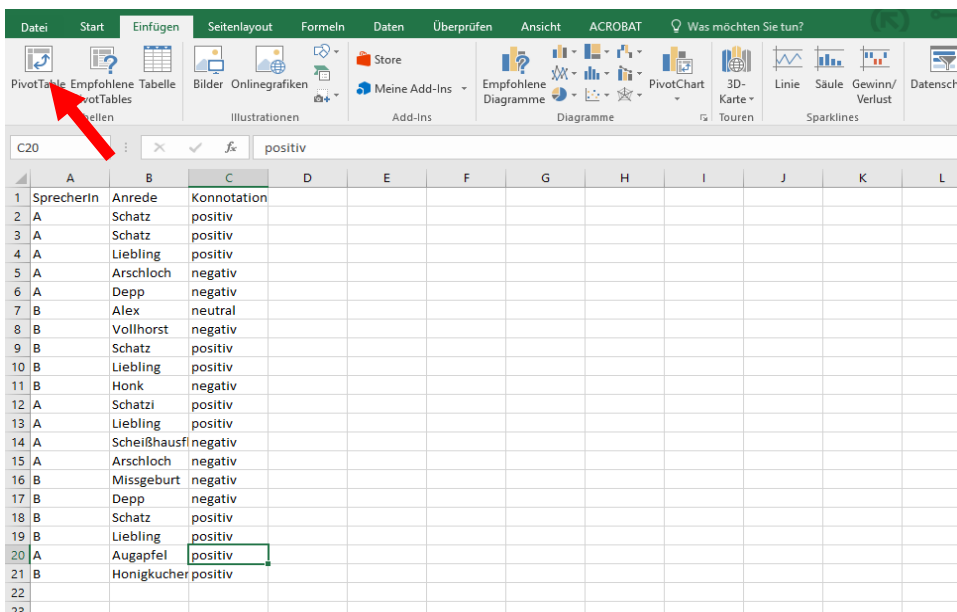
- **entweder** Strg+Alt+V und wählen im sich öffnenden Dialog „Werte“
- **oder** benutzen die Maus: Rechtsklick > Inhalte einfügen > Werte einfügen (Screenshot unten).

| | A | B | C | D |
|----|-----------|--------------|-------------|-------------|
| 1 | Sprecher1 | Anrede | Konnotation | Zufallszahl |
| 2 | A | Schatz | positiv | 0.83 |
| 3 | B | Missgeburt | negativ | 0.83 |
| 4 | B | Liebling | positiv | 0.05 |
| 5 | B | Liebling | positiv | 0.3 |
| 6 | A | Liebling | positiv | 0.26 |
| 7 | A | Liebling | positiv | 0.21 |
| 8 | A | Augapfel | positiv | 0.27 |
| 9 | B | Honk | negativ | 0.60 |
| 10 | B | Honigkuchen | positiv | 0.38 |
| 11 | A | Schatzi | positiv | 0.20 |
| 12 | B | Alex | neutral | 0.45 |
| 13 | B | Vollhorst | negativ | 0.83 |
| 14 | A | Depp | negativ | 0.47 |
| 15 | B | Depp | negativ | 0.1 |
| 16 | A | Arschloch | negativ | 0.50 |
| 17 | A | Scheißhaufen | negativ | 0.26 |
| 18 | B | Schatz | positiv | 0.43 |
| 19 | B | Schatz | positiv | 0.56 |
| 20 | A | Schatz | positiv | 0.46 |
| 21 | A | Arschloch | negativ | 0.14 |
| 22 | | | | |
| 23 | | | | |

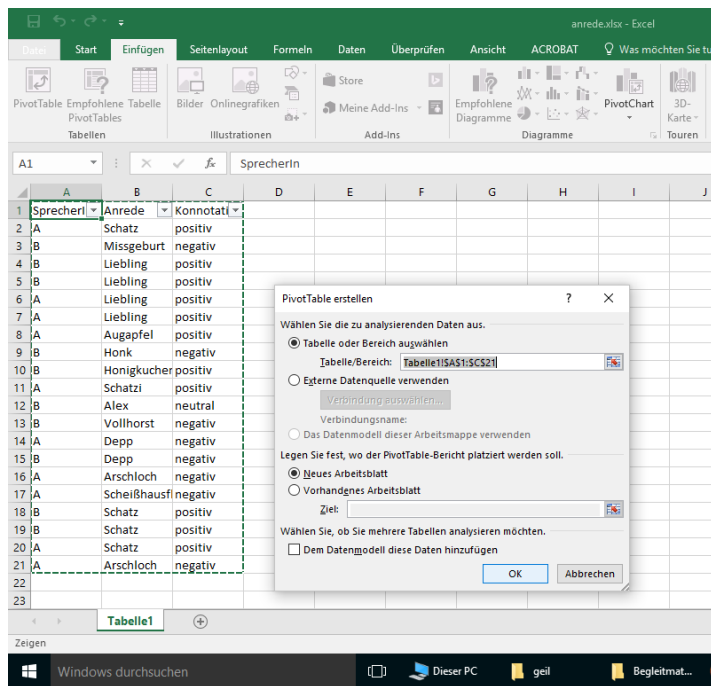
Beide Wege führen zum gleichen Ziel. Nicht wundern: Die Zufallsdaten werden noch einmal neu generiert und ändern sich daher. Aber wenn Sie jetzt in die Spalte mit den eingefügten Daten klicken und in die Formelzeile oben schauen, werden Sie sehen, dass dort nicht mehr =ZUFALLSZAHL() steht, sondern vielmehr der eingefügte Wert. Dieser bleibt nun konstant.

1.2 Tabellen

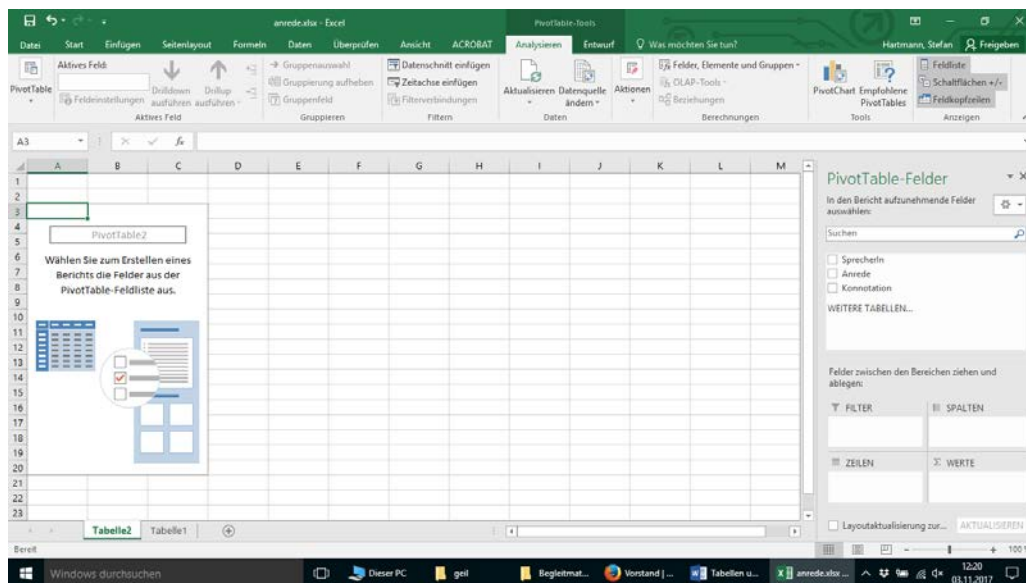
Schauen wir uns nun an, wie man in Excel einfache Tabellen erstellen kann. Angenommen z.B., wir wollen wissen, wie häufig in unseren Daten welches Lexem gebraucht wird, unabhängig von der Sprecherin oder dem Sprecher. Dafür bietet sich die *PivotTable*-Funktion an, die oben links im Reiter „Einfügen“ versteckt ist:



Wenn nicht gerade mehrere Zellen markiert sind, wählt PivotTable als den Bereich, dessen Daten in die Tabelle eingehen sollen, standardmäßig alle ausgefüllten Zellen, wie der folgende Screenshot zeigt:



Wenn wir nun einfach OK klicken, sehen wir noch keine Tabelle, sondern lediglich eine Art Grundgerüst dafür:



Wir können nun eine Tabelle generieren, indem wir die Variablen, die wir im Fenster mittig rechts sehen, in die kleinen Fenster unten ziehen. Wenn wir z.B. eine Tabelle im folgenden Format wollen...

| | Sprecher/in A | Sprecher/in B |
|---------|---------------|---------------|
| negativ | 4 | 4 |
| neutral | 0 | 1 |
| positiv | 6 | 5 |

... dann müssen wir zunächst „SprecherIn“ ins Fenster „Spalten“ ziehen und „Konnotation“ ins Fenster „Zeilen“ - denn in der Tabelle sollen ja A und B eigene Spalten haben, während

die unterschiedlichen Konnotationen in den Zeilen erscheinen. Wenn wir das erledigt haben, fehlt allerdings noch das, was uns eigentlich interessiert, nämlich die eigentliche Auswertung. Der entscheidende Schritt ist daher nun, „Konnotation“ auch noch ins Fensterchen „Werte“ zu ziehen - denn wir wollen ja, dass Excel die einzelnen Variablenausprägungen in der Spalte „Konnotation“ auszählt. Am Ende steht eine Tabelle, die ungefähr so aussieht wie die oben gezeigte.

| Anzahl von Konnotation | Spaltenbeschriftungen | | |
|------------------------|-----------------------|----|----------------|
| Zeilenbeschriftungen | A | B | Gesamtergebnis |
| negativ | 4 | 4 | 8 |
| neutral | 1 | | 1 |
| positiv | 6 | 5 | 11 |
| Gesamtergebnis | 10 | 10 | 20 |

1.3 Grafiken

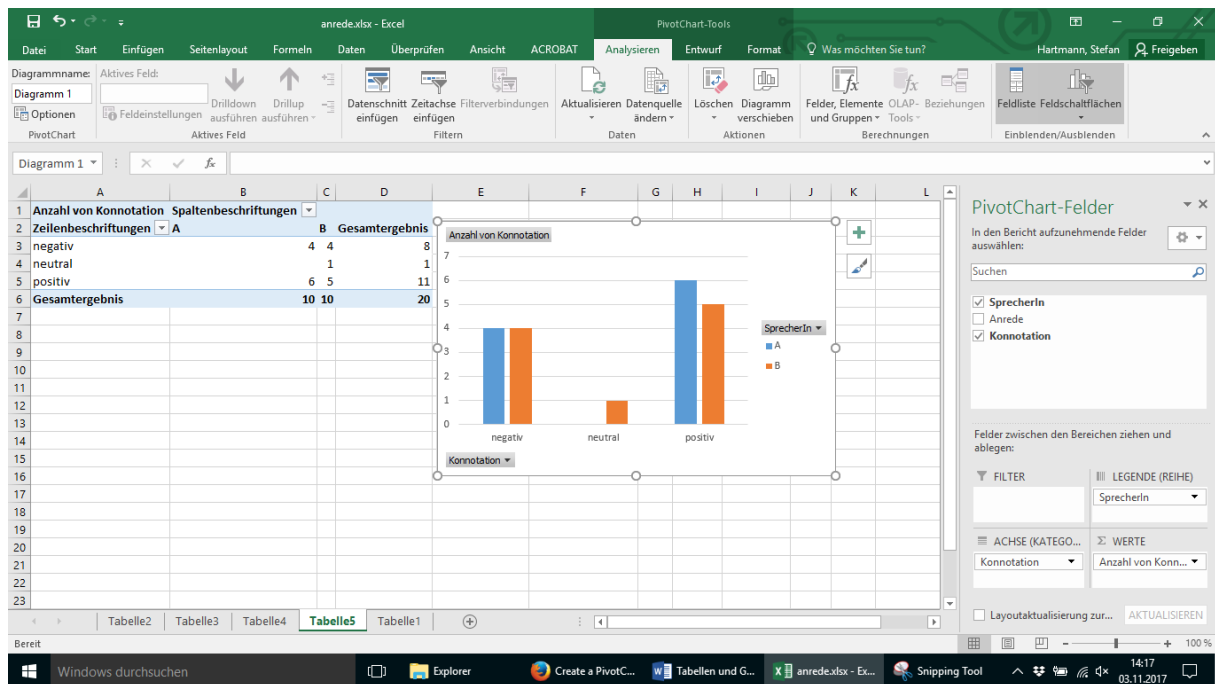
Außer Tabellen können wir in Excel auch einfache Grafiken erstellen.² Hier gibt es einen wichtigen **Unterschied zwischen der PC- und der Mac-Version** von Excel: Unter Windows gibt es eine Funktion „PivotChart“, die im Grunde genauso funktioniert wie die Pivot-Tabelle, die wir gerade erstellt haben. Bei der Mac-Version hingegen fehlt diese Funktion. Wenn wir hier mit solchen Rohdaten³ arbeiten, wie wir sie in unsrer Beispieltabelle vorfinden, müssen wir beim Mac den Umweg über die Pivot-Table-Funktion wählen, d.h. wir müssen erst eine Pivot-Tabelle erstellen und diese dann als Input für unsere Grafik(en) verwenden.

Werfen wir aber zunächst einen Blick auf die derzeit nur unter Windows verfügbare Variante. Hier können wir, genau wie bei der PivotTable-Funktion oben, einfach auf „PivotChart“ klicken; auf dem neuen Arbeitsblatt, wo das Diagramm erstellt wird, sehen wir dann wieder ein leeres „Grundgerüst“, das wir mit Hilfe der Fenster auf der rechten Seite füllen müssen. Die Fensterchen heißen hier etwas anders als bei der PivotTable-Funktion, funktionieren aber im Grunde genauso.

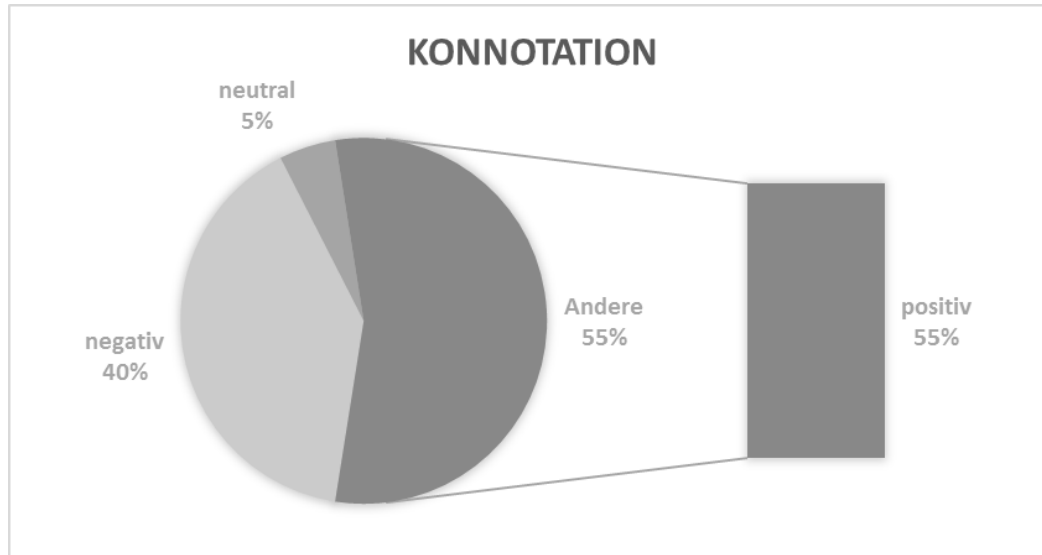
Wenn wir zum Beispiel die Variable *SprecherIn* ins Fenster LEGENDE (REIHE) ziehen und *Konnotation* sowohl in ACHSE (KATEGORIE) als auch in WERTE, dann erhalten wir ein Balkendiagramm, das wie die oben erstellte Tabelle die Spalte „Konnotation“ nach Sprecher/in getrennt auswertet.

² Für komplexere Visualisierungen würde ich jedoch eher R empfehlen. Dies auch deshalb, weil der Export von Excel-Diagrammen in gängige Grafikformate in hochauflösendem Format gelinde gesagt schwierig ist. Vgl. dazu das entsprechende Tutorial „Schnelleinstieg in R“ mit ausführlichem Visualisierungskapitel.

³ Damit meine ich, dass die Daten noch nicht in irgendeiner Weise aggregiert/ausgewertet sind. Manchmal arbeiten wir auch mit bereits aggregierten Daten, z.B. Frequenzlisten. (Außerdem sind die Daten schon deshalb roh, weil sie Lexeme wie *Scheißhausfliege* oder *Missgeburt* enthalten...)



Wenn wir statt einem Barplot lieber eine andere Visualisierungsmöglichkeit nutzen wollen, dann können wir ganz oben auf „PivotChart Tools“ klicken und im Reiter „Entwurf“ die Option „Diagrammtyp ändern“ auswählen. „Star Trek“-Fans können die Daten damit sogar in Form des Raumschiffs Enterprise darstellen:



Allerdings macht diese Visualisierungsvariante nicht wirklich viel Sinn. (Ein solches Tortendiagramm ist z.B. für Parteien gedacht, wo die große „Sonstige“-Kategorie dann noch einmal aufgesplittet werden kann.) Generell sollte man sich vor dem sogenannten „Overplotting“ hüten und nicht durch unnötige visuelle Spielereien von dem ablenken, was die Grafik eigentlich zeigen soll. Etwas ausführlicher gehe ich darauf im Tutorial „Schnelleinstieg in R“ ein.

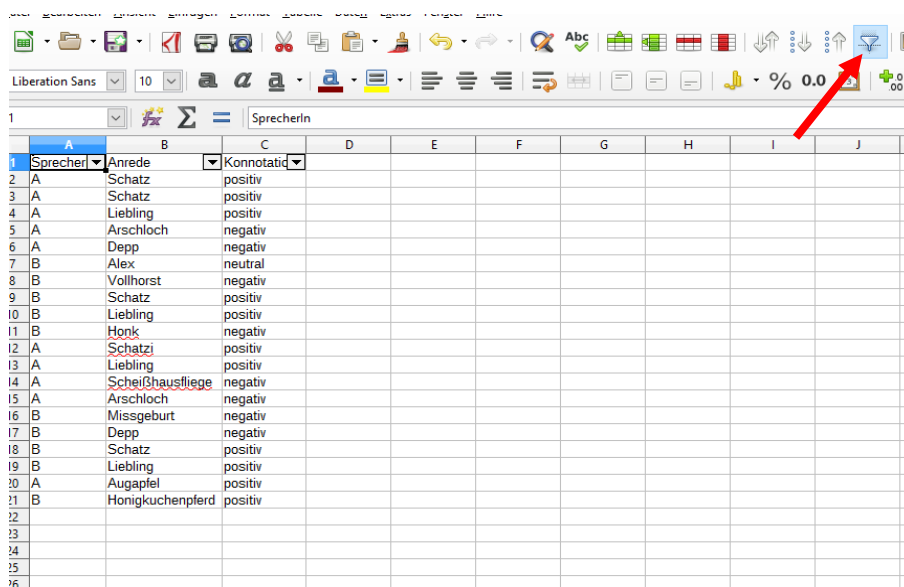
Unter **Mac** kann man die PivotChart-Funktion, wie gesagt, nicht verwenden, aber man kann einfach so wie oben dargestellt eine Pivot-Tabelle erstellen. Mit Cmd+A kann man dann die gesamte Pivot-Tabelle markieren und unter Einfügen > Diagramme dann ein Diagramm auswählen, das man zur Visualisierung nutzen möchte.

2. Tabellen und Grafiken mit Calc

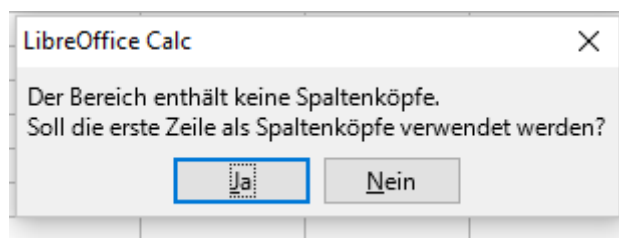
Anstelle des kommerziellen Microsoft Excel kann man auch das kostenlose Programm LibreOffice Calc erwähnen. Im Tutorial „01-Grundlegendes“ habe ich bereits erklärt, dass es zwar z.T. einen geringeren Funktionsumfang als Excel hat, dafür aber auch einige Vorzüge. So gestaltet sich das Einlesen der Daten intuitiver, da Calc gleich beim Öffnen einer Datei nach Encoding, Trennzeichen usw. fragt. Sind die Daten erst eingelesen, halten sich die Unterschiede in den wesentlichen Funktionen in Grenzen. Daher ist der Text im Folgenden z.T. schamlos aus dem Excel-Tutorial oben übernommen.

2.1 Daten sortieren und filtern

In sehr vielen Fällen kann es nützlich sein, Daten zu sortieren (z.B. alphabetisch) oder zu filtern. Um von diesen Möglichkeiten Gebrauch machen zu können, können wir den Befehl „AutoFilter“ verwenden (unter Daten > AutoFilter; alternativ einfach auf das Filtersymbol in der Leiste ganz oben klicken).



Voraussetzung dafür, dass das funktioniert, ist, dass in der ersten Zeile die Spaltennamen stehen (was hier der Fall ist). Bevor wir auf „Filtern“ klicken, stellen wir sicher, dass irgendeine Zelle in der ersten Zeile ausgewählt ist (wie es oben im Screenshot der Fall ist). Wenn folgende Meldung kommt:



bestätigen wir mit „Ja“ - denn wie eben gesagt, wollen wir ja, dass die erste Zeile als Überschrift (= Spaltenköpfe) behandelt wird.

Sobald wir auf „Filtern“ geklickt haben, erscheint neben jedem Spaltennamen ein kleines Pfeilsymbol (siehe Screenshot oben).

Sehen wir uns nun zunächst an, wie man Daten **sortiert**. Angenommen, wir wollen nur Daten von Sprecherin A und nur positiv konnotierte oder neutrale Anreden sehen. Dann können wir

zunächst auf das Pfeilchen neben *SprecherIn* klicken und das Häkchen bei „B“ entfernen. Nun werden uns nur noch Daten von Sprecherin A angezeigt. Wichtig: Die Daten von Sprecherin B sind nach wie vor in der Datei vorhanden, sie sind nur ausgeblendet! Wenn wir nun auf das Häkchen neben „Konnotation“ klicken, sehen wir, dass uns nur zwei Kategorien angezeigt werden: „positiv“ und „negativ“. Die Kategorie „neutral“ wird nicht angezeigt, weil sie nur in den ausgeblendeten Daten vorkommt! Wenn wir nun nur die positiv konnotierten Daten sehen wollen, genügt es, das Häkchen bei „negativ“ zu entfernen.

Trial and error

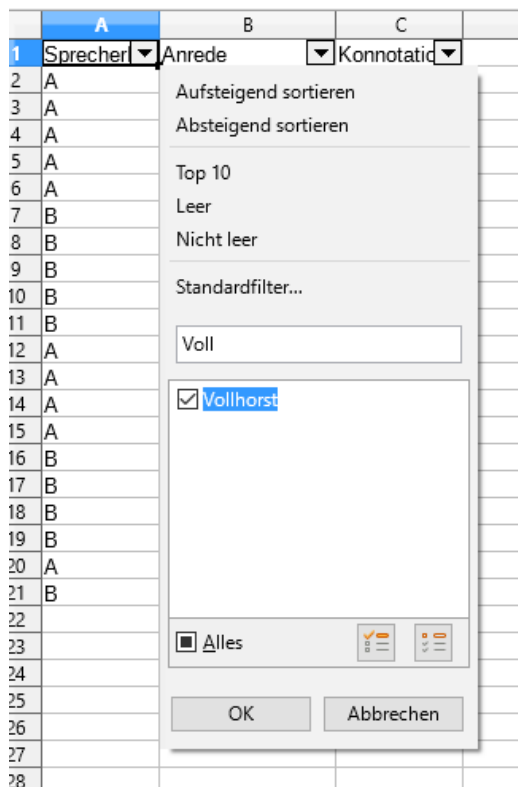
Was passiert, wenn wir nun „B“ wieder einblenden?

Wenn wir nun „B“ wieder einblenden, sehen wir nach wie vor nur „positive“ Belege - die Kategorie „neutral“ müssen wir manuell wieder einblenden, wenn wir alle positiven und neutralen Daten von A und B haben wollen!

Stolperfalle!

Wenn Sie mit mehreren Filtern arbeiten und einen der Filter verändern, kontrollieren Sie bei Bedarf die anderen Filter, um zu verhindern, dass Sie (z.B. bei der Annotation) Daten übersehen, die für Sie relevant sind.

Gerade bei größeren Datensätzen kann es jedoch sinnvoll sein, nicht jede einzelne Variablenausprägung durch Setzen oder Entfernen von Häkchen auszuwählen, sondern stattdessen mit dem Textfilter zu arbeiten, mit dem ich beispielsweise alle Anreden suchen kann, die mit *Voll-* beginnen:



Wie Sie bereits in den obigen Screenshots gesehen haben, haben Sie darüber hinaus die Möglichkeit, die Daten aufsteigend bzw. absteigend zu sortieren. Das kann insbesondere dann sinnvoll sein, wenn Sie Daten haben, in denen einige Types mehrfach vorkommen und bei denen Sie sicher sein können, dass alle Tokens desselben Types dieselbe Annotation bekommen: Dann können Sie einfach das erste Token annotieren und die Annotation dann in alle anderen Zeilen copy&pasten (oder sogar schon vorher alle Zeilen, in die die Annotation kommen soll, markieren, und dann mit Alt+Eingabetaste die eingegebene Annotation auf alle markierten Zeilen übertragen).

Ein Beispiel: Wir wollen unsere positiven und negativen Anreden, bei denen es sich zum großen Teil um Metaphern (oder z.T. auch Metonymien sowie Grenzgängern zwischen beiden) handelt, auf ihren Quellbereich hin annotieren. Dann ist es sinnvoll, die Daten alphabetisch zu sortieren, sodass wir die Tokens, die zum gleichen Type gehören, zusammen annotieren können:

| | A | B | C | D |
|----|----------|------------------|-------------|--------------|
| 1 | Sprecher | Anrede | Konnotation | Quellbereich |
| 2 | B | Alex | neutral | NA |
| 3 | A | Arschloch | negativ | Körperteil |
| 4 | A | Arschloch | negativ | |
| 5 | A | Augapfel | positiv | |
| 6 | A | Depp | negativ | |
| 7 | B | Depp | negativ | |
| 8 | B | Honigkuchenpferd | positiv | |
| 9 | B | Honk | negativ | |
| 10 | A | Liebling | positiv | |
| 11 | B | Liebling | positiv | |
| 12 | A | Liebling | positiv | |
| 13 | B | Liebling | positiv | |
| 14 | B | Missgeburt | negativ | |
| 15 | A | Schatz | positiv | |
| 16 | A | Schatz | positiv | |
| 17 | B | Schatz | positiv | |
| 18 | B | Schatz | positiv | |
| 19 | A | Schatzi | positiv | |
| 20 | A | Scheißhausfliege | negativ | |
| 21 | B | Vollhorst | negativ | |
| 22 | | | | |
| 23 | | | | |
| 24 | | | | |

Im obigen Screenshot sind die Zellen D3 bis D5 markiert. Es gibt zwei Möglichkeiten, das zu erreichen:

- Zelle D3 markieren und bei gedrückter Umschalt-Taste⁴ die Pfeil-nach-unten-Taste drücken.
- Zelle D3 markieren und das kleine Quadrat, das Sie unten rechts an der Markierung sehen, nach unten ziehen. Gerade wenn Sie sehr viele Zellen markieren möchten, kann das manchmal der effizientere Weg sein,

Die Eingabe erfolgt nach wie vor in der ersten Zeile. Wenn Sie nach der Eingabe nur die Eingabetaste drücken, wird Ihre Annotation nicht in die übrigen markierten Spalten übertragen - das geschieht nur, wenn Sie Alt+Eingabetaste drücken.

⁴ Die Umschalt- oder Shift-Taste ist die Taste ganz links, die Sie sonst benutzen, um Großbuchstaben zu schreiben.

Alternativ können wir auch *Körperteil* in Zelle D3 schreiben, dann Zelle D3 markieren und die Annotation durch Ziehen am kleinen Quadrat unten rechts an der Markierung auf die beiden folgenden Zeilen übertragen.

Ganz nebenbei sehen Sie im obigen Screenshot übrigens auch eine Konvention, wie man mit Daten umgehen kann, auf die die Annotationskriterien nicht zutreffen und die man deshalb bei der Annotation auslassen möchte: Da *Alex* ein Eigenname ist und daher keinen Quellbereich hat, habe ich ihn mit NA (= Not Available) annotiert.

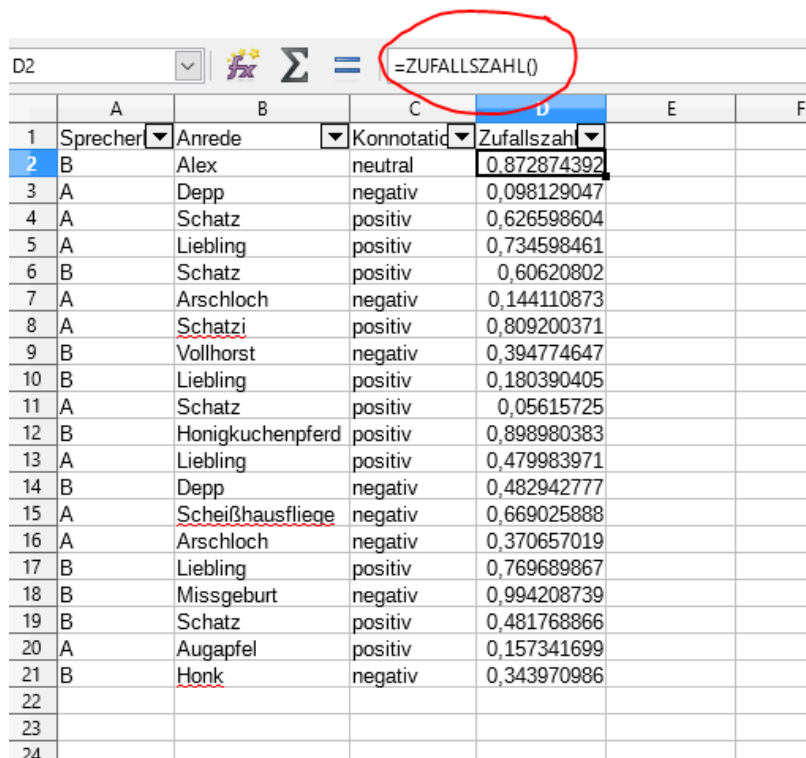
2.2 Eine Spalte mit Zufallszahlen hinzufügen

Manchmal wollen wir aber die Daten gerade *nicht* nach Kriterien sortieren, die sich aus den Daten selbst ergeben, sondern vielmehr zufällig. Ein Beispiel: Wenn wir diachrone Daten annotieren und die Hypothese haben, dass diachron eine bestimmte Veränderung auftreten wird, dann wäre es sehr problematisch, die Daten nach Jahren sortiert zu annotieren, denn es könnte sein, dass wir bei der Annotation unsere Hypothese zu sehr im Hinterkopf haben und in Zweifelsfällen daher - bewusst oder unbewusst - eher dazu neigen, so zu entscheiden, dass das Ergebnis unserer Hypothese entspricht. In solchen Fällen kann es sinnvoll sein, eine Spalte mit Zufallszahlen hinzuzufügen. Die Formel

=ZUFALLSZAHN()

generiert eine Zufallszahl zwischen 0 und 1. Wenn wir diese Formel in der ersten Zeile anwenden und dann in alle anderen Zeilen übertragen (wiederum indem wir am kleinen Quadrat unten rechts an der Markierung ziehen), und dann die Daten nach dieser neuen Spalte sortieren, dann sind die Daten in einer zufälligen Reihenfolge.

Allerdings: Die Zufallszahlen werden jedes Mal neu generiert. Wenn Sie das Dokument schließen und wieder öffnen, dann haben Sie neue Zufallszahlen und damit auch eine neue Reihenfolge. In der Regel wollen Sie das nicht. Aber auch hier hilft ein Trick. Wenn Sie in eine beliebige Zelle der „Zufallszahl“-Spalte klicken, sehen Sie, dass oben in der Formelspalte noch unsere Formel steht:



The screenshot shows a spreadsheet with the following data:

| | A | B | C | D | E | F |
|----|----------|------------------|-------------|-------------|---|---|
| 1 | Sprecher | Anrede | Konnotation | Zufallszahl | | |
| 2 | B | Alex | neutral | 0,872874392 | | |
| 3 | A | Depp | negativ | 0,098129047 | | |
| 4 | A | Schatz | positiv | 0,626598604 | | |
| 5 | A | Liebling | positiv | 0,734598461 | | |
| 6 | B | Schatz | positiv | 0,60620802 | | |
| 7 | A | Arschloch | negativ | 0,144110873 | | |
| 8 | A | Schatzi | positiv | 0,809200371 | | |
| 9 | B | Vollhorst | negativ | 0,394774647 | | |
| 10 | B | Liebling | positiv | 0,180390405 | | |
| 11 | A | Schatz | positiv | 0,05615725 | | |
| 12 | B | Honigkuchenpferd | positiv | 0,898980383 | | |
| 13 | A | Liebling | positiv | 0,479983971 | | |
| 14 | B | Depp | negativ | 0,482942777 | | |
| 15 | A | Scheißhausfliege | negativ | 0,669025888 | | |
| 16 | A | Arschloch | negativ | 0,370657019 | | |
| 17 | B | Liebling | positiv | 0,769689867 | | |
| 18 | B | Missgeburt | negativ | 0,994208739 | | |
| 19 | B | Schatz | positiv | 0,481768866 | | |
| 20 | A | Augapfel | positiv | 0,157341699 | | |
| 21 | B | Honk | negativ | 0,343970986 | | |
| 22 | | | | | | |
| 23 | | | | | | |
| 24 | | | | | | |

Das bedeutet, dass hinter den Werten, die wir in der Spalte sehen, immer noch der Zufallsgenerator steht. Jedoch können wir die Kopierfunktion nutzen, um in eine andere (oder auch in dieselbe) Spalte nicht die Formel einzufügen, die wir kopiert haben, sondern die Werte, die sie hervorbringt. Dafür markieren wir zunächst die Zellen, die wir kopieren wollen, und kopieren sie mit Strg + C oder Rechtsklick > Kopieren in die Zwischenablage. Als nächstes wählen wir dort, wo wir die Ergebnisse einfügen wollen, Rechtsklick > Inhalte einfügen > Zahlen.

Nicht wundern: Die Zufallsdaten werden noch einmal neu generiert und ändern sich daher. Aber wenn Sie jetzt in die Spalte mit den eingefügten Daten klicken und in die Formelzeile oben schauen, werden Sie sehen, dass dort nicht mehr =ZUFALLSZAHN() steht, sondern vielmehr der eingefügte Wert. Dieser bleibt nun konstant.

| D3 | | | | | 0,317697992548347 |
|----|----------|------------------|-------------|-------------|-------------------|
| | A | B | C | D | E |
| 1 | Sprecher | Anrede | Konnotation | Zufallszahl | |
| 2 | B | Alex | neutral | 0,771565246 | |
| 3 | A | Depp | negativ | 0,317697993 | |
| 4 | A | Schatz | positiv | 0,613151161 | |
| 5 | A | Liebling | positiv | 0,383934893 | |
| 6 | B | Schatz | positiv | 0,56168872 | |
| 7 | A | Arschloch | negativ | 0,723988297 | |
| 8 | A | Schatzi | positiv | 0,118855628 | |
| 9 | B | Vollhorst | negativ | 0,167487916 | |
| 10 | B | Liebling | positiv | 0,049650678 | |
| 11 | A | Schatz | positiv | 0,880217979 | |
| 12 | B | Honigkuchenpferd | positiv | 0,199994411 | |
| 13 | A | Liebling | positiv | 0,114915723 | |
| 14 | B | Depp | negativ | 0,949771109 | |
| 15 | A | Scheißhausfliege | negativ | 0,496876429 | |
| 16 | A | Arschloch | negativ | 0,030143714 | |
| 17 | B | Liebling | positiv | 0,537840461 | |
| 18 | B | Missgeburt | negativ | 0,901447655 | |
| 19 | B | Schatz | positiv | 0,649989137 | |
| 20 | A | Augapfel | positiv | 0,88435774 | |
| 21 | B | Honk | negativ | 0,757655892 | |
| 22 | | | | | |
| 23 | | | | | |

2.2 Tabellen

Schauen wir uns nun an, wie man in Calc einfache Tabellen erstellen kann. Angenommen z.B., wir wollen wissen, wie häufig in unseren Daten welches Lexem gebraucht wird, unabhängig von der Sprecherin oder dem Sprecher. Dafür bietet sich die *PivotTable*-Funktion an, die man unter Einfügen > PivotTable findet. Wenn nicht gerade mehrere Zellen markiert sind, wählt PivotTable als den Bereich, dessen Daten in die Tabelle eingehen sollen, standardmäßig alle ausgefüllten Zellen, wie der folgende Screenshot zeigt:

A1:D21

| | A | B | C | D | E | F | G | H | I |
|----|------------|------------------|-------------|-------------|---|---|---|---|---|
| 1 | SprecherIn | Anrede | Konnotation | Zufallszahl | | | | | |
| 2 | B | Alex | neutral | 0,771565246 | | | | | |
| 3 | A | Depp | negativ | 0,317697993 | | | | | |
| 4 | A | Schatz | positiv | 0,613151161 | | | | | |
| 5 | A | Liebling | positiv | 0,383934893 | | | | | |
| 6 | B | Schatz | positiv | 0,56168872 | | | | | |
| 7 | A | Arschloch | negativ | 0,723988297 | | | | | |
| 8 | A | Schatzi | positiv | 0,118855628 | | | | | |
| 9 | B | Vollhorst | negativ | 0,167487916 | | | | | |
| 10 | B | Liebling | positiv | 0,049650678 | | | | | |
| 11 | A | Schatz | positiv | 0,880217979 | | | | | |
| 12 | B | Honigkuchenpferd | positiv | 0,199994411 | | | | | |
| 13 | A | Liebling | positiv | 0,114915723 | | | | | |
| 14 | B | Depp | negativ | 0,949771109 | | | | | |
| 15 | A | Scheißhaustiege | negativ | 0,496876429 | | | | | |
| 16 | A | Arschloch | negativ | 0,030143714 | | | | | |
| 17 | B | Liebling | positiv | 0,537840461 | | | | | |
| 18 | B | Missgeburt | negativ | 0,901447655 | | | | | |
| 19 | B | Schatz | positiv | 0,649989137 | | | | | |
| 20 | A | Augapfel | positiv | 0,88435774 | | | | | |
| 21 | B | Honk | negativ | 0,757655892 | | | | | |
| 22 | | | | | | | | | |
| 23 | | | | | | | | | |
| 24 | | | | | | | | | |
| 25 | | | | | | | | | |
| 26 | | | | | | | | | |

Quelle auswählen

Auswahl

☒ Aktuelle Auswahl

☐ Benannter Bereich:

☐ In LibreOffice angemeldete Datenquelle

Hilfe OK Abbrechen

Es öffnet sich nun ein Fenster „PivotTabellen-Layout“, in dem wir festlegen müssen, was in den Daten in den Zeilen und Spalten erscheinen soll:

Pivot-Tabellen Layout

Seitenfelder:

Spaltenfelder:

Daten

Zeilenfelder:

Datenfelder:

Verfügbare Felder:

SprecherIn

Anrede

Konnotation

Zufallszahl

Ziehen Sie die Elemente in die gewünschte Position

☐ Optionen

☐ Quelle und Zielbereich

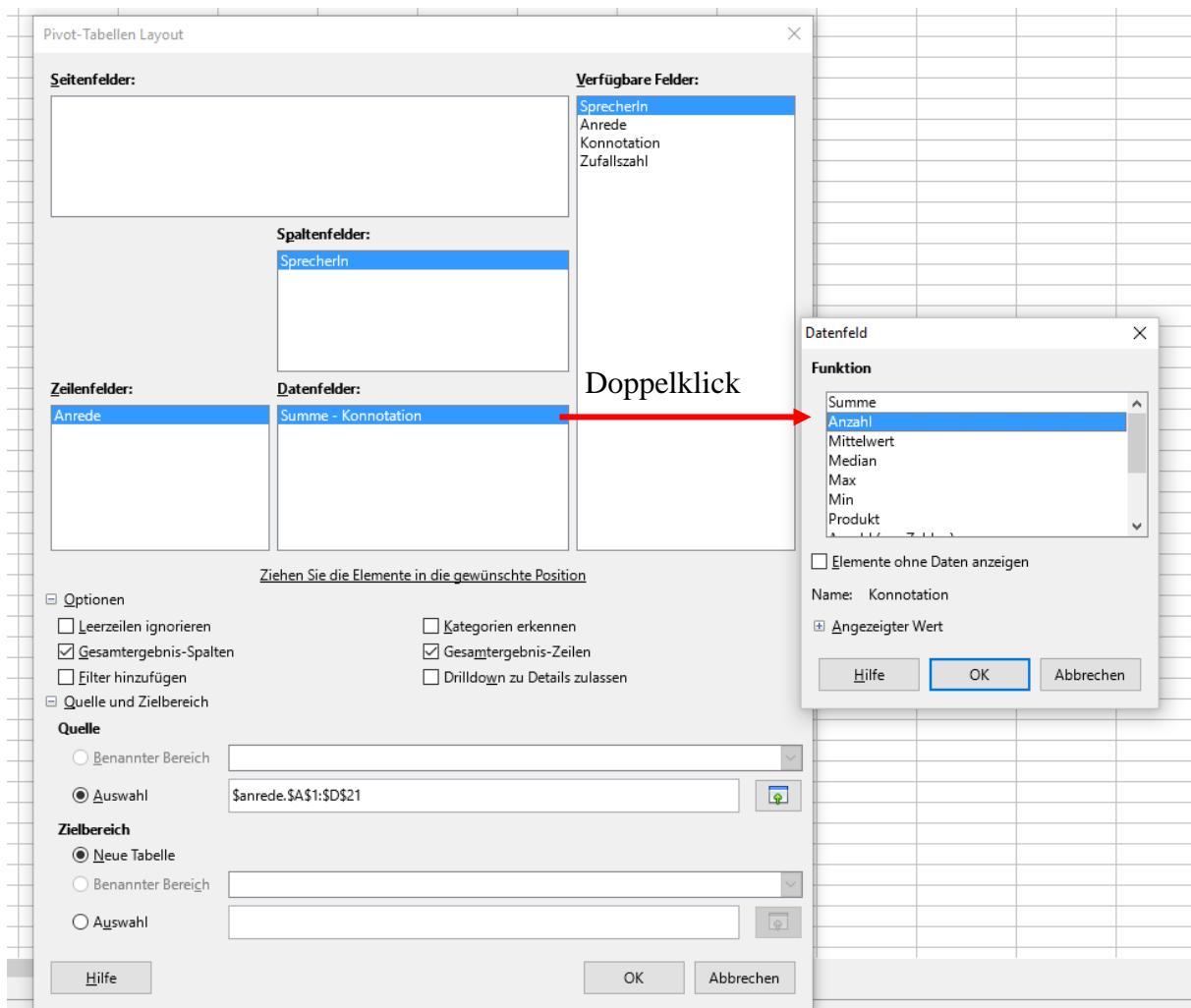
Hilfe OK Abbrechen

Mit Hilfe dieser „Bausteine“ können wir nun eine Tabelle generieren, indem wir die Variablen, die im Fenster „Verfügbare Felder“ stehen, in die kleinen Fenster unten ziehen. Wenn wir z.B. eine Tabelle im folgenden Format wollen.

| | Sprecher/in A | Sprecher/in B |
|---------|---------------|---------------|
| negativ | 4 | 4 |

| | | |
|---------|---|---|
| neutral | 0 | 1 |
| positiv | 6 | 5 |

... dann müssen wir zunächst „SprecherIn“ ins Fenster „Spaltenfelder“ ziehen und „Konnotation“ ins Fenster „Zeilenfelder“ - denn in der Tabelle sollen ja Sprecher/in A und B eigene Spalten haben, während die unterschiedlichen Konnotationen in den Zeilen erscheinen. Wenn wir das erledigt haben, fehlt allerdings noch das, was uns eigentlich interessiert, nämlich die eigentliche Auswertung. Der entscheidende Schritt ist daher nun, „Konnotation“ auch noch ins Fensterchen „Datenfelder“ zu ziehen - denn wir wollen ja, dass Calc die einzelnen Variablenausprägungen in der Spalte „Konnotation“ auszählt. Defaultmäßig zählt Calc allerdings die Variablenausprägungen nicht aus, sondern bildet eine Summe, wie Sie im Screenshot unten sehen können: Im Feld „Datenfelder“ steht „Summe - Konnotation“. Eine Summe bilden kann man aber natürlich nur von numerischen Variablen. „Positiv/negativ/neutral“ sind nicht numerisch. Damit Calc keine Summe zu bilden versucht (und kläglich scheitert), sondern auszählt, müssen wir auf „Summe - Konnotation“ im Fenster „Datenfelder“ doppelklicken und anschließend die Option „Anzahl“ wählen.



Am Ende steht eine Tabelle, die ungefähr so aussieht wie die oben gezeigte:

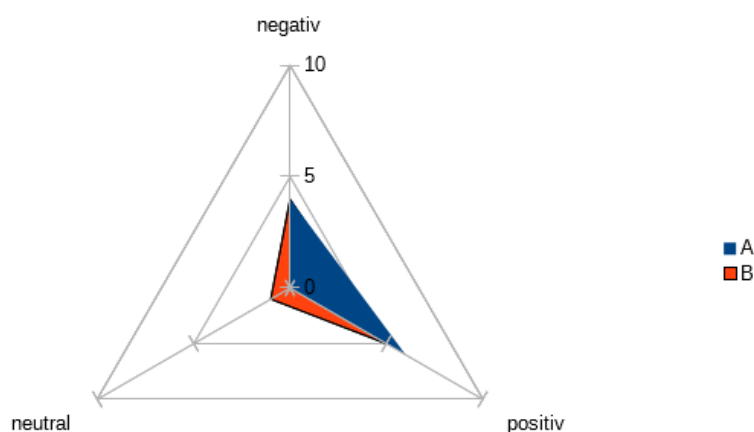
| Anzahl - Konnotation | SprecherIn | | |
|------------------------|------------|-----------|-----------------|
| Konnotation | A | B | Gesamt Ergebnis |
| negativ | 4 | 4 | 8 |
| neutral | | 1 | 1 |
| positiv | 6 | 5 | 11 |
| Gesamt Ergebnis | 10 | 10 | 20 |

2.3 Grafiken

Außer Tabellen können wir in Calc auch einfache Grafiken erstellen.⁵ Anders als bei Excel für Windows gibt es bei Calc keine PivotChart-Funktion, sodass wir zur Diagrammerstellung aus unseren Rohdaten den Umweg über die Pivot-Tabelle nehmen müssen, die wir gerade erstellt haben. Hier markieren wir zunächst die für uns relevanten Spalten, wobei wir das „Gesamt Ergebnis“, wie Calc es in etwas normwidriger Graphie nennt, außen vor lassen:

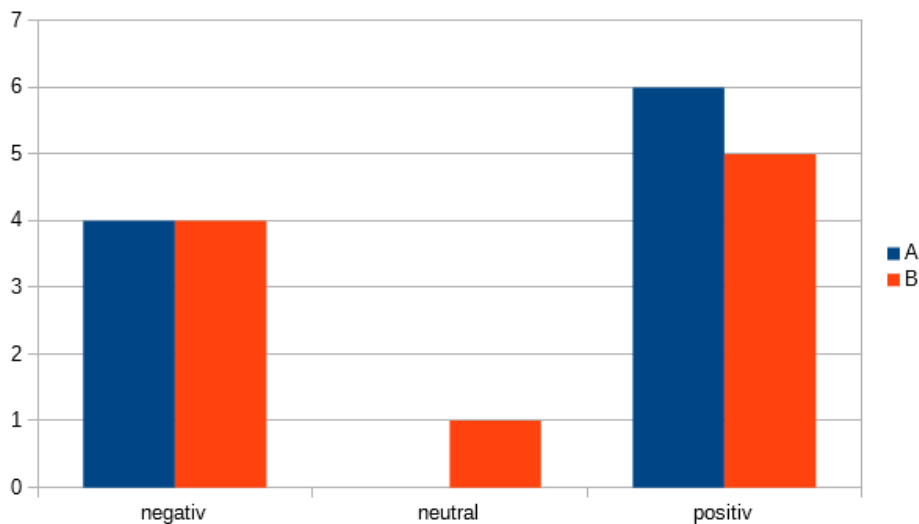
| | A | B | C | D |
|---|------------------------|------------|-----------|-----------------|
| 1 | Anzahl - Konnotation | SprecherIn | | |
| 2 | Konnotation | A | B | Gesamt Ergebnis |
| 3 | negativ | 4 | 4 | 8 |
| 4 | neutral | | 1 | 1 |
| 5 | positiv | 6 | 5 | 11 |
| 6 | Gesamt Ergebnis | 10 | 10 | 20 |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |

Nun können wir unter Einfügen > Diagramm zunächst einen Diagrammtyp wählen. Dabei sollte man natürlich ein möglichst leicht interpretierbares Format wählen und nicht unbedingt so etwas, auch wenn es ästhetisch aussehen mag:

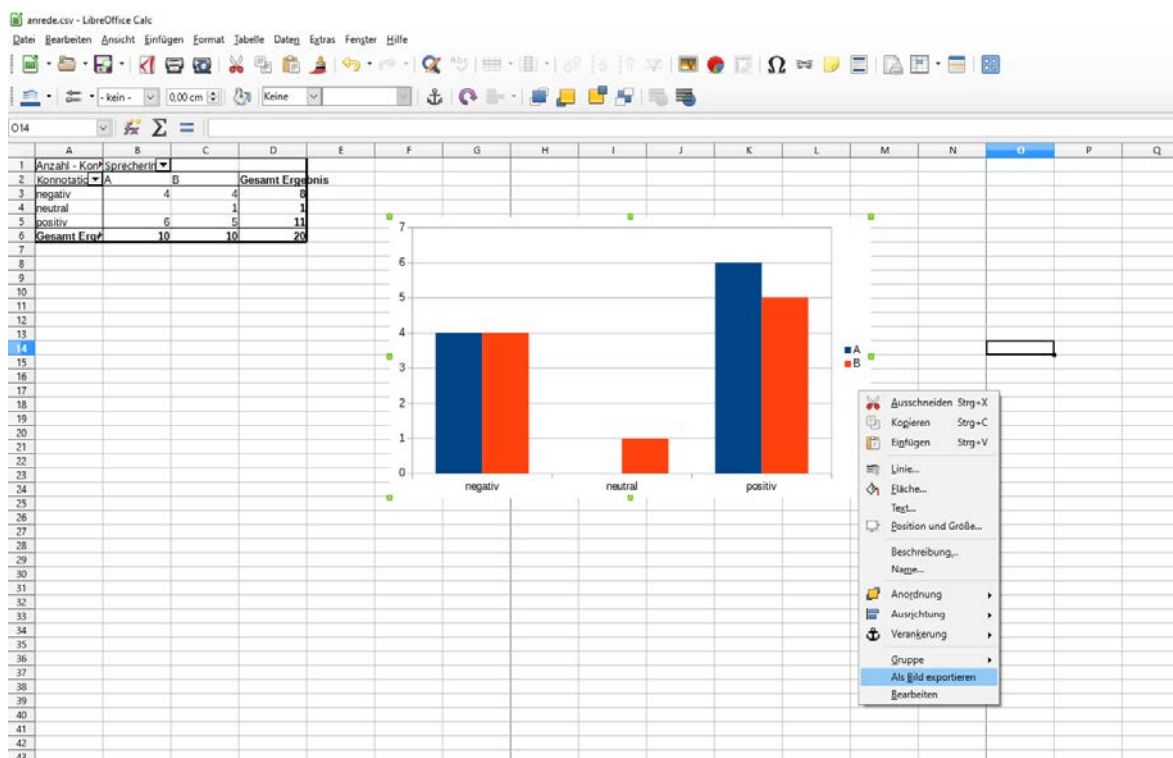


Geeigneter ist für unsere Zwecke z.B. ein klassisches Säulendiagramm, was von Calc auch tatsächlich als erstes vorgeschlagen wird.

⁵ Für komplexere Visualisierungen würde ich jedoch eher R empfehlen.



Wenn man einen Balken anklickt, kann man unter Format > Auswahl formatieren (oder einfach auf den Balken doppelklicken) z.B. im Reiter „Fläche“ die Farbe des Balkens verändern, und auch sonst kann man das Diagramm mit dieser Funktion den eigenen Bedürfnissen anpassen. Mit Rechtsklick > „Als Bild exportieren“ kann man das Diagramm dann auch als Grafikdatei exportieren, wobei diese Option nur erscheint, wenn man auch wirklich das Diagramm als Ganzes angewählt hat und nicht nur einzelne Elemente davon:



Erfreulicherweise kann man die Diagramme hier auch in den Vektorgrafik-Formaten SVG und WMF abspeichern, sodass man sie ohne Qualitätsverlust z.B. in Word einfügen kann (aber Achtung: Word für Windows unterstützt zwar ab der Version 2016 das SVG-Format, Word für Mac jedoch nicht; das Windows-eigene unterstützen aber beide Word-Versionen.)