

Erste Schritte und erste Datenexplorationen in CODAP (Arbeitsblatt 1)

Beim Öffnen von CODAP sieht die Arbeitsoberfläche wie folgt aus:

The screenshot shows the CODAP interface with a table titled 'JIM_Paderborn' containing 215 cases. The table has the following columns: Index, Geschlecht, Alter, Klasse, Anzahl_Apps, Anzahl_LineProfile, Anzahl_Kontakte, Anzahl_Impressions, Musik_hoeren, Kino_gehen, Lesen_gelungen, Lesen_Lesungen, Lesen_geschrieben, Lesen_Lesungen, Nutzen_Handy, Schauliste, and Schauliste_Videos. The table is displayed in a grid view with a blue header and a light blue background.

| Index | Geschlecht | Alter | Klasse | Anzahl_Apps | Anzahl_LineProfile | Anzahl_Kontakte | Anzahl_Impressions | Musik_hoeren | Kino_gehen | Lesen_gelungen | Lesen_Lesungen | Lesen_geschrieben | Lesen_Lesungen | Nutzen_Handy | Schauliste | Schauliste_Videos |
|-------|------------|-------|---------|-------------|--------------------|-----------------|--------------------|--------------|------------|----------------|----------------|-------------------|----------------|--------------|------------|-------------------|
| 1 | maennl. | 12 | 6 | 92 | 9 | 88 | 415 | 1_taepli. | 5_einm. | 5_einm. | 5_einm. | 5_einm. | 6_selte. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 2 | maennl. | 12 | 6 | 34 | 5 | 120 | 500 | 1_taepli. | 6_selte. | 3_einm. | 6_selte. | 4_einm. | 7_nie | 1_taepli. | 7_nie | 1_taepli. |
| 3 | weiblich | 18 | 11 (Q1) | 13 | 1 | 70 | 969 | 1_taepli. | 5_einm. | 1_taepli. | 7_nie | 1_taepli. | 7_nie | 1_taepli. | 2_mehr. | 2_mehr. |
| 4 | maennl. | 13 | 7 | 40 | 15 | 1265 | 700 | 1_taepli. | 6_selte. | 7_nie | 7_nie | 7_nie | 7_nie | 1_taepli. | 1_1_tae. | 1_taepli. |
| 5 | maennl. | 17 | 11 (Q1) | 104 | 10 | 314 | 1000 | 1_taepli. | 5_einm. | 2_mehr. | 4_einm. | 6_selte. | 6_selte. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 6 | maennl. | 17 | 11 (Q1) | 83 | 12 | 177 | 982 | 1_taepli. | 6_selte. | 4_einm. | 4_einm. | 5_einm. | 5_einm. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 7 | maennl. | 17 | 11 (Q1) | 170 | 40 | 271 | 4356 | 1_taepli. | 5_einm. | 1_taepli. | 2_mehr. | 6_selte. | 4_einm. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 8 | maennl. | 16 | 10 (EF) | 70 | 9 | 239 | 3012 | 2_mehr. | 5_einm. | 2_mehr. | 6_selte. | 2_mehr. | 6_selte. | 1_taepli. | 4_einm. | 2_mehr. |
| 9 | maennl. | 13 | 8 | 35 | 6 | 270 | 560 | 1_taepli. | 5_einm. | 1_taepli. | 2_mehr. | 2_mehr. | 4_einm. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 10 | maennl. | 14 | 8 | 10 | 5 | 87 | 629 | 2_mehr. | 5_einm. | 5_einm. | 6_selte. | 5_einm. | 6_selte. | 1_taepli. | 1_1_tae. | 2_mehr. |
| 11 | maennl. | 18 | 12 (Q2) | 110 | 25 | 311 | 4000 | 1_taepli. | 6_selte. | 6_selte. | 7_nie | 7_nie | 5_einm. | 1_taepli. | 2_mehr. | 1_taepli. |
| 12 | weiblich | 17 | 11 (Q1) | 128 | 1 | 46 | 1466 | 2_mehr. | 6_selte. | 1_taepli. | 1_taepli. | 5_einm. | 1_taepli. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 13 | weiblich | 18 | 12 (Q2) | 6 | 2 | 153 | 2059 | 6_selte. | 6_selte. | 2_mehr. | 7_nie | 7_nie | 7_nie | 1_taepli. | 7_nie | 2_mehr. |
| 14 | maennl. | 17 | 11 (Q1) | 18 | 6 | 200 | 500 | 1_taepli. | 5_einm. | 5_einm. | 6_selte. | 5_einm. | 6_selte. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 15 | weiblich | 18 | 12 (Q2) | 45 | 4 | 190 | 511 | 3_einm. | 6_selte. | 4_einm. | 2_mehr. | 6_selte. | 7_nie | 1_taepli. | 3_einm. | 3_einm. |
| 16 | weiblich | 14 | 9 | 51 | 5 | 170 | 561 | 1_taepli. | 6_selte. | 3_einm. | 6_selte. | 4_einm. | 7_nie | 1_taepli. | 2_mehr. | 4_einm. |
| 17 | maennl. | 18 | 11 (Q1) | 26 | 3 | 310 | 200 | 1_taepli. | 6_selte. | 2_mehr. | 2_mehr. | 4_einm. | 5_einm. | 1_taepli. | 5_einm. | 1_taepli. |
| 18 | maennl. | 16 | 10 (EF) | 62 | 1 | 38 | 1242 | 1_taepli. | 6_selte. | 7_nie | 7_nie | 7_nie | 7_nie | 1_taepli. | 2_mehr. | 3_einm. |
| 19 | maennl. | 16 | 11 (Q1) | 26 | 5 | 154 | 834 | 1_taepli. | 5_einm. | 5_einm. | 1_taepli. | 5_einm. | 1_taepli. | 1_taepli. | 1_1_tae. | 1_taepli. |
| 20 | maennl. | 17 | 11 (Q1) | 75 | 2 | 150 | 1000 | 1_taepli. | 6_selte. | 6_selte. | 7_nie | 7_nie | 7_nie | 1_taepli. | 2_mehr. | 4_einm. |

Die grundlegenden Funktionalitäten finden sich in der Arbeitsleiste links oben:



Zunächst kommen wir mit den Elementen **Graph** und **Text** aus.

Ihr erster Arbeitsauftrag (ca. 10 Minuten)

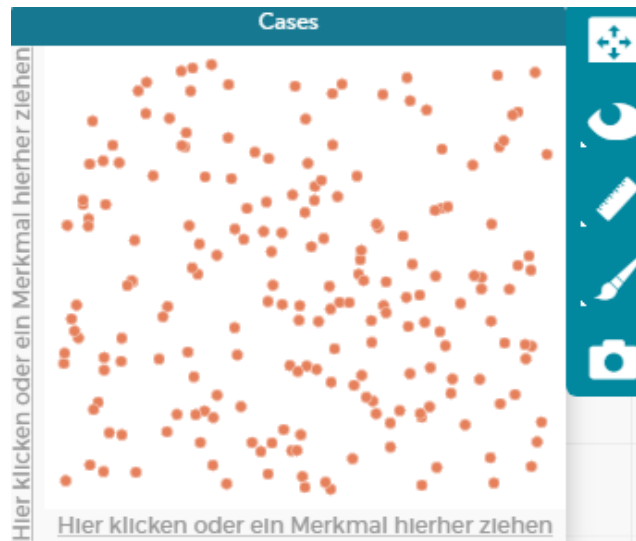
Öffnen Sie den reduzierten JIM-PB-Datensatz unter Nutzung des folgenden Links (Nutzen Sie als Browser google Chrome).

Link: <https://tinyurl.com/JIM-DZLM>

Machen Sie dann die Schritte, die auf den folgenden Seiten erklärt sind, in CODAP nach.

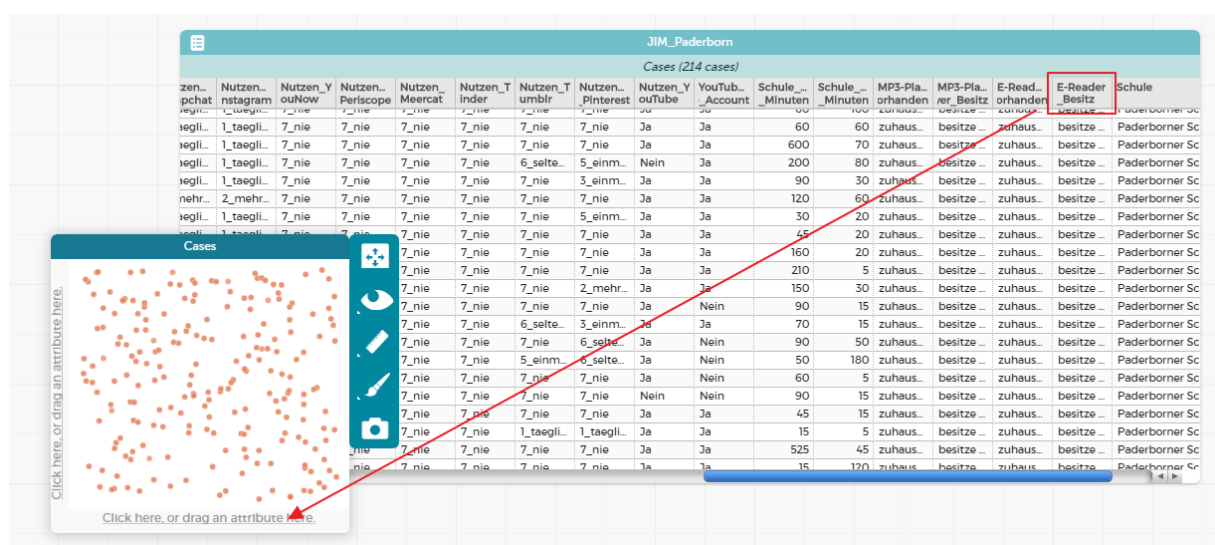
Graph

Klickt man in der Arbeitsleiste auf „Graph“, so öffnet sich ein Graph, in dem die Daten aus der Tabelle als Punkte dargestellt sind. Jeder Punkt entspricht einem Schüler bzw. einer Schülerin und einer Zeile in der Datentabelle.



Nun kann man den Graph nutzen, um die Daten nach selbst-gewählten Fragestellungen zu explorieren:

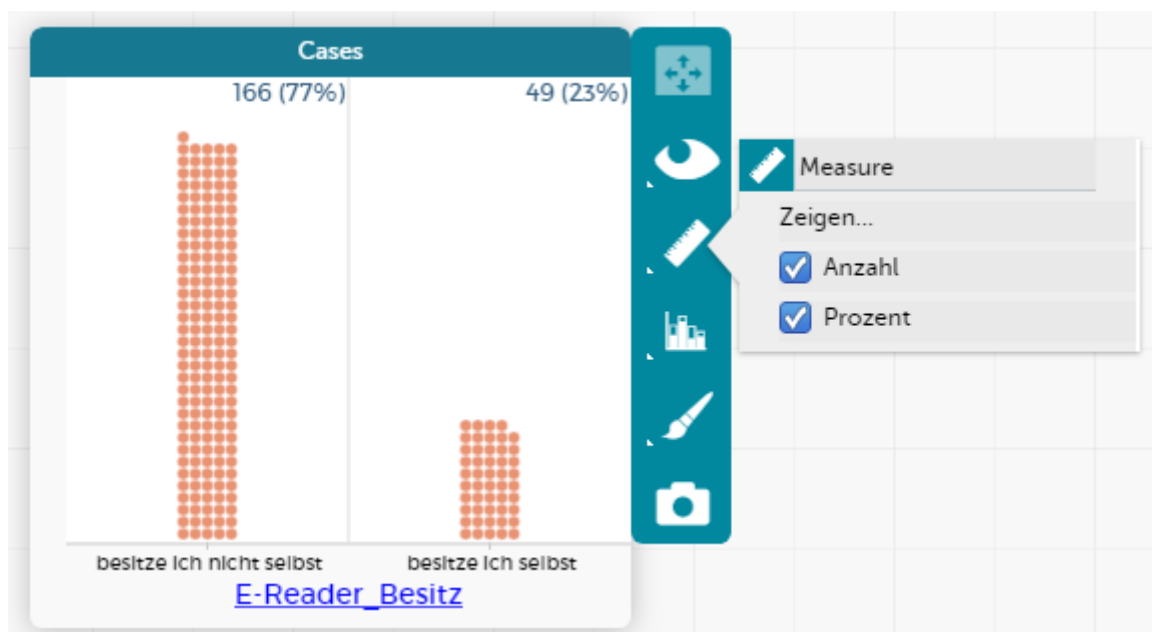
Will man beispielsweise einen Eindruck über die **Verteilung des kategorialen Merkmals „E-Reader_Besitz“** bekommen, so kann man das Merkmal in der Tabelle auswählen und auf die gewünschte Achse des Graphen aufrufen.



Man erhält dann ein gestapeltes Punktdiagramm zur Verteilung des Merkmals in CODAP:

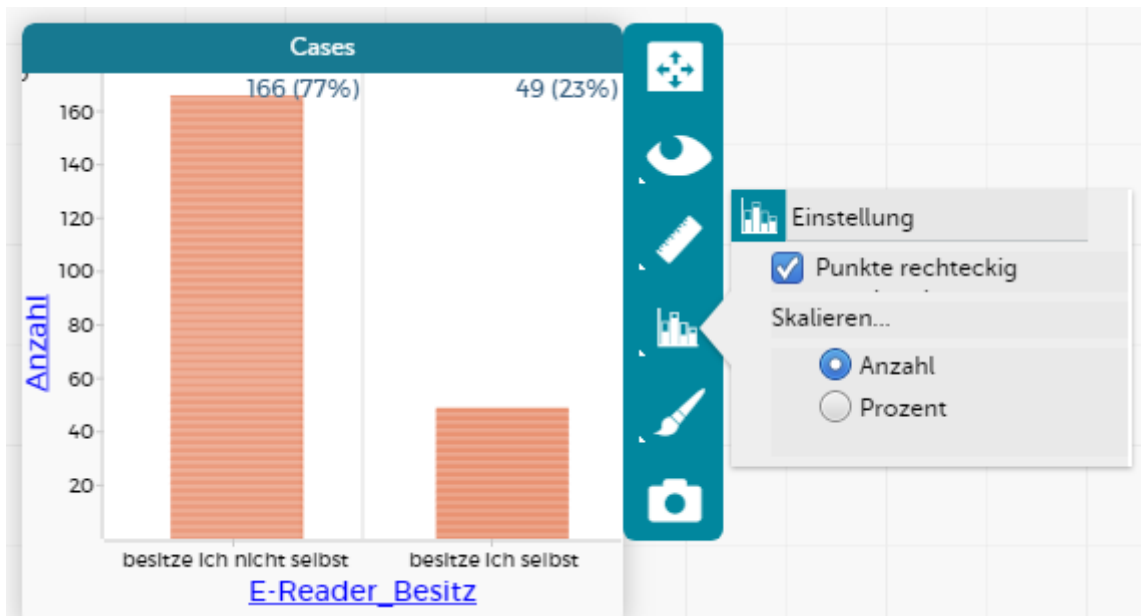


Qualitativ lässt sich auf dem ersten Blick sagen, dass ein Großteil der Schülerinnen und Schüler keinen E-Reader besitzt. Die Leiste rechts neben dem Graph bietet dann weitere Optionen an (wie z.B. **Anzeigen von absoluten und relativen Häufigkeiten, Verändern der Darstellungsform, Verbergen von Fällen** etc.). Die entsprechenden **absoluten und relativen Häufigkeiten** kann man von CODAP beispielsweise bestimmen lassen, indem man das Lineal und dann die Optionen „Anzahl“ und „Prozent“ auswählt.

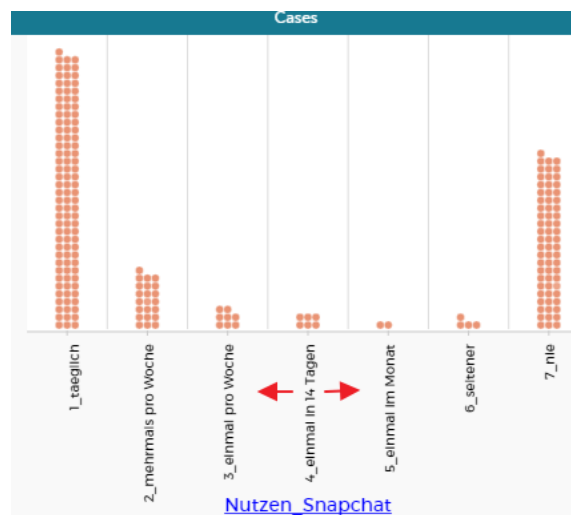


Hier lässt sich nun der Graphik entnehmen, dass 23% der befragten Schülerinnen und Schüler einen E-Reader und 77% der Befragten keinen E-Reader besitzen.

Außerdem lässt sich die **Darstellungsform verändern**, indem man die Punkte rechteckig zu einem konventionellen **Säulendiagramm** verschmelzen lässt.



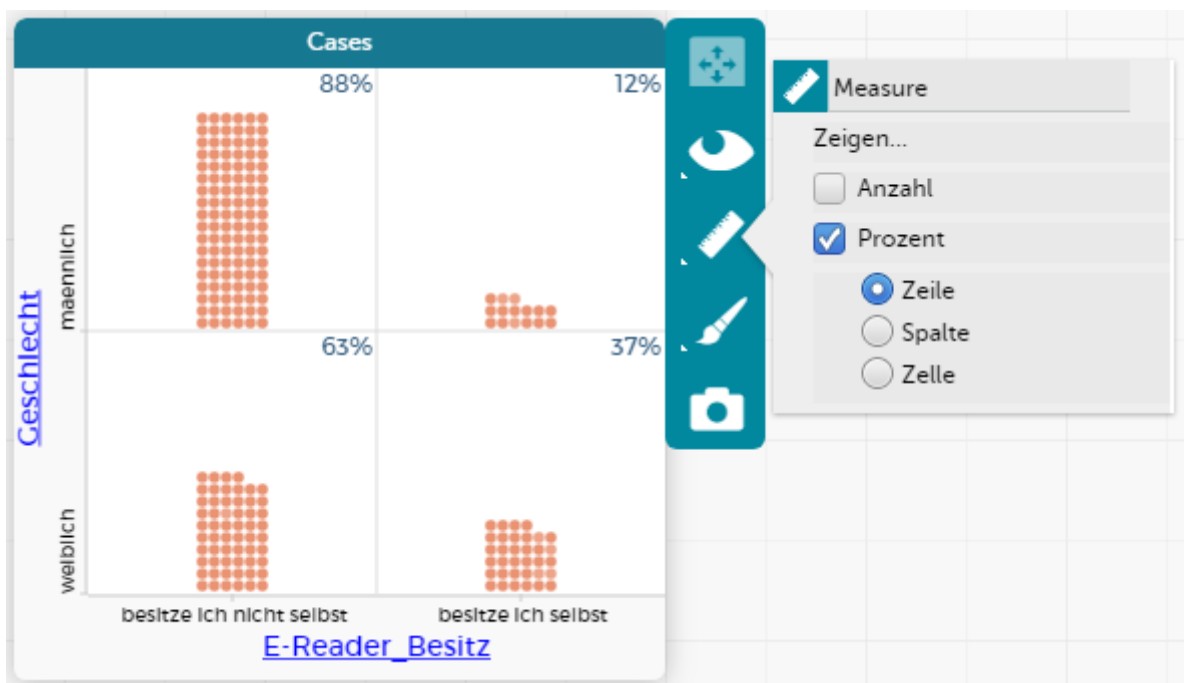
Bei **kategorialen Merkmalen**, die **mehr als zwei Ausprägungen** aufweisen, wie zum Beispiel das Merkmal „Nutzen Snapchat“, kann es sinnvoll sein (falls im Datensatz noch nicht geschehen), die Ausprägungen (täglich, etc.) zu ordnen. Dazu können die Ausprägungen z.B. aufsteigend nach der Nutzungsintensität angeordnet werden.



Will man den **Zusammenhang zwischen zwei kategorialen Merkmalen** untersuchen und die Frage beantworten, ob tendenziell eher die Jungen oder die Mädchen einen E-Reader besitzen, so lässt sich ein weiteres Merkmal, beispielsweise das Merkmal „Geschlecht“, auf die y-Achse ziehen.

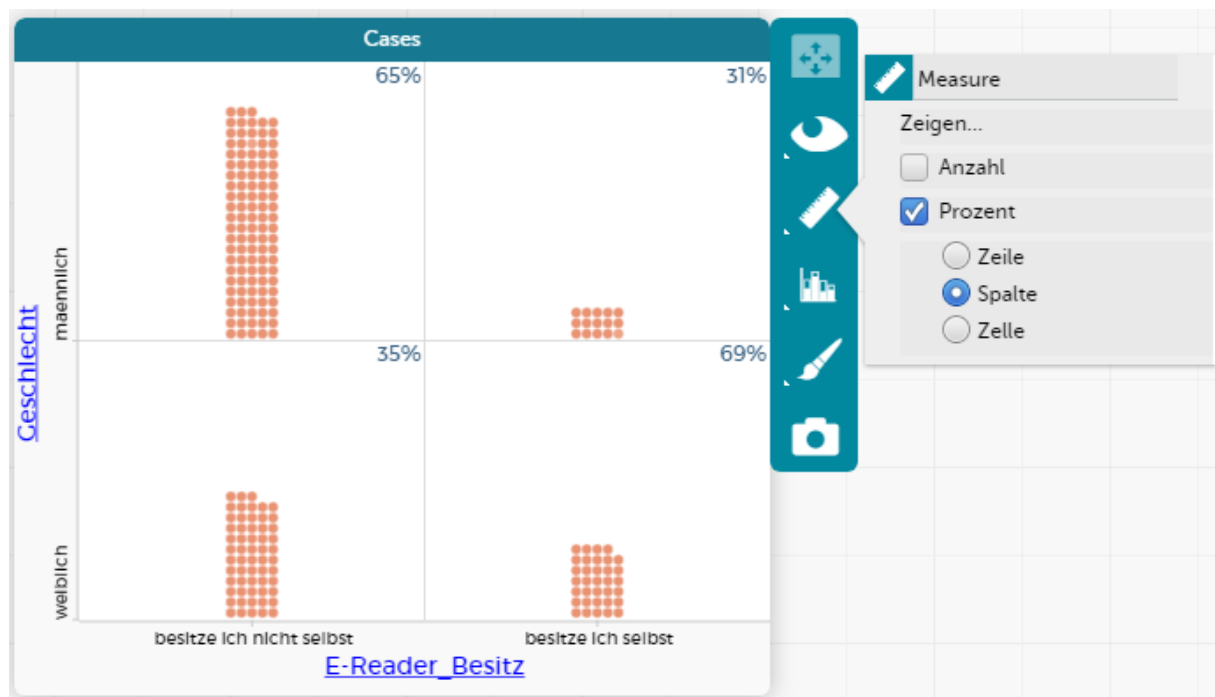


Auch hier bietet die Arbeitsleiste rechts neben dem Graphen weitere Optionen an (wie z.B. Verändern der Darstellungsform, Anzeigen von absoluten und relativen Häufigkeiten, Verbergen von Fällen etc.). Die Auswahl von **verschiedenen Prozentsen (Zeile, Spalte, Zelle)** ermöglicht verschiedene Auswertungsperspektiven auf die Fragestellung. In unserem Fall und bezüglich unserer Fragestellung (Besitzen tendenziell eher die Jungen oder die Mädchen einen E-Reader?) wollen wir in der Teilgruppe der Jungen und in der Teilgruppe der Mädchen den Anteil der E-Reader-Besitzer bestimmen. Dazu wählen wir **Zeilenprozente** in CODAP aus.



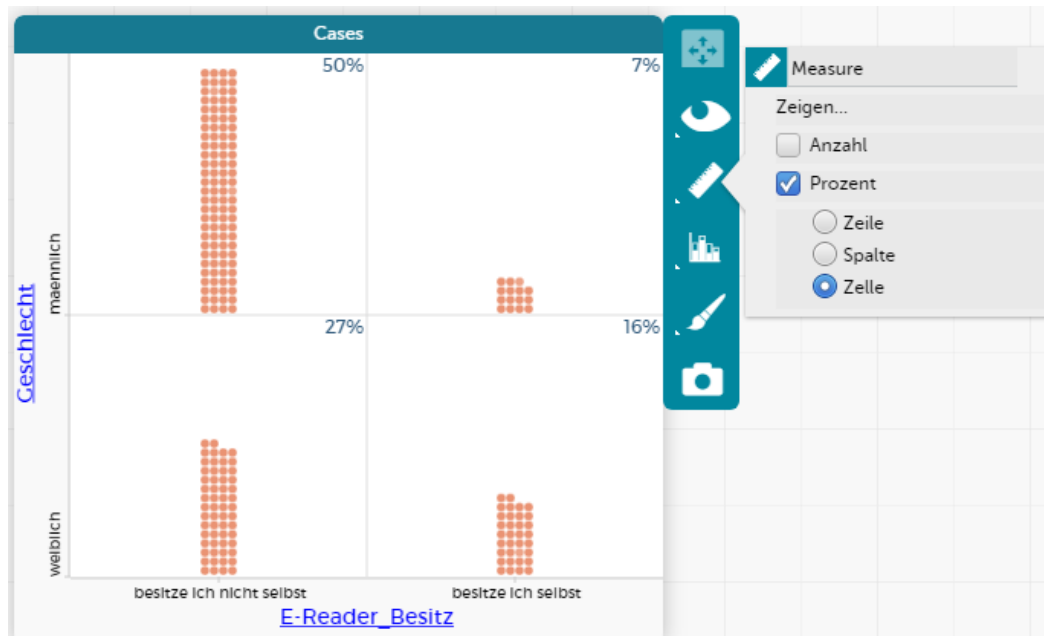
Mithilfe dieser Graphik können wir unsere Ausgangsfrage beantworten: Die Mädchen besitzen tendenziell eher einen E-Reader als die Jungen, denn 37% der Mädchen und nur 12% der Jungen besitzen einen E-Reader (bzw. 88% der Jungen und 63% der Mädchen besitzen keinen E-Reader).

Anstelle von **Zeilenprozenten** könnte man in CODAP auch **Spaltenprozente** berechnen lassen – so können wir der folgenden Graphik entnehmen, dass in der Gruppe der „Nicht-E-Reader-Besitzer“ 65% männlich und 35% weiblich sind.

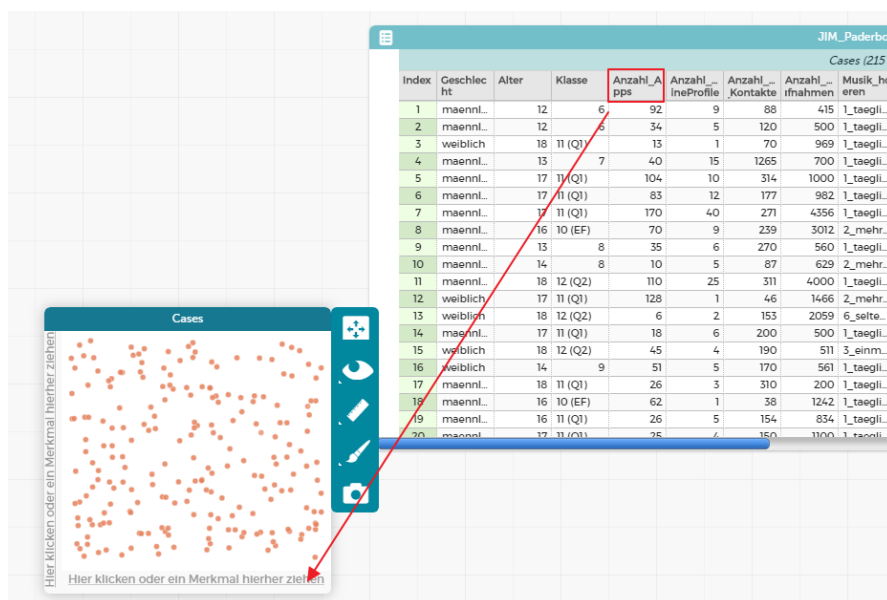


Oder anders ausgedrückt: 65% von den „Nicht-E-Reader-Besitzern“ sind Schüler und 35% von der „Nicht-E-Reader-Besitzern“ sind Schülerinnen. Das scheint unsere oben getroffene Aussage zu untermauern - es gilt hier aber zu bedenken, dass bei dieser spaltenweisen Auswertung der höhere Anteil an Schülern in der „Nicht-E-Reader-Besitzer“-Gruppe auch dadurch bedingt sein kann, dass die Schüler gegenüber den Schülerinnen in dem vorliegenden Datensatz ein Übergewicht haben (123 Schüler vs. 92 Schülerinnen). Für die Auswertung unserer Fragestellung (Besitzen tendenziell eher die Jungen oder die Mädchen einen E-Reader?) eignen sich daher in unserem Falle die **Zeilenprozente**.

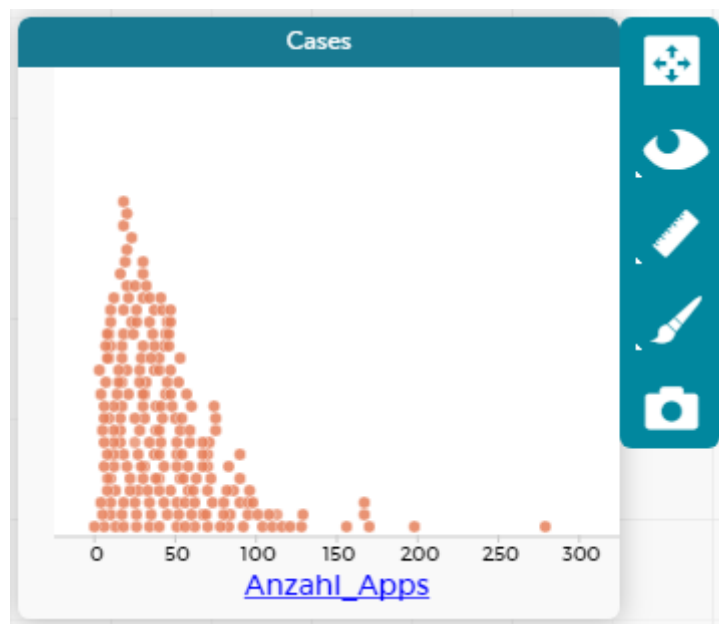
Schließlich lassen sich in CODAP auch **Zellenprozente** berechnen. So lässt sich mit der folgenden Darstellung aussagen, dass 50% der Befragten männlich und „Nicht-E-Reader-Besitzer“ und 27% der Befragten weiblich und „Nicht-E-Reader-Besitzer“ sind. Dementsprechend sind 7% der Befragten männlich und „E-Reader-Besitzer“ sowie 16% der in diesem Datensatz Befragten weiblich und „E-Reader-Besitzer“.



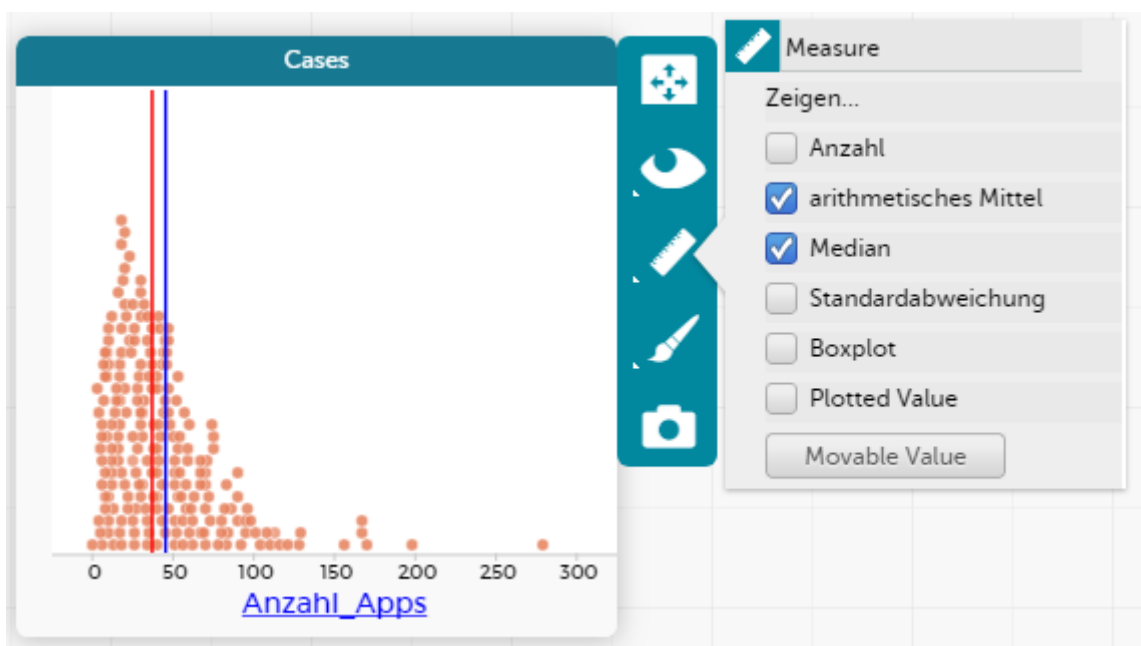
Nun haben wir bis jetzt ausschließlich Verteilungen eines **kategorialen Merkmals** betrachtet. Der JIM-PB-Datensatz enthält aber auch **numerische Merkmale**. Stellen wir uns vor, wir interessieren uns für die **Verteilung des numerischen Merkmals** Anzahl_Apps im JIM-Datensatz. Ähnlich wie bei kategorialen Merkmalen, wählt man das entsprechende numerische Merkmal aus und zieht es auf die x-Achse:



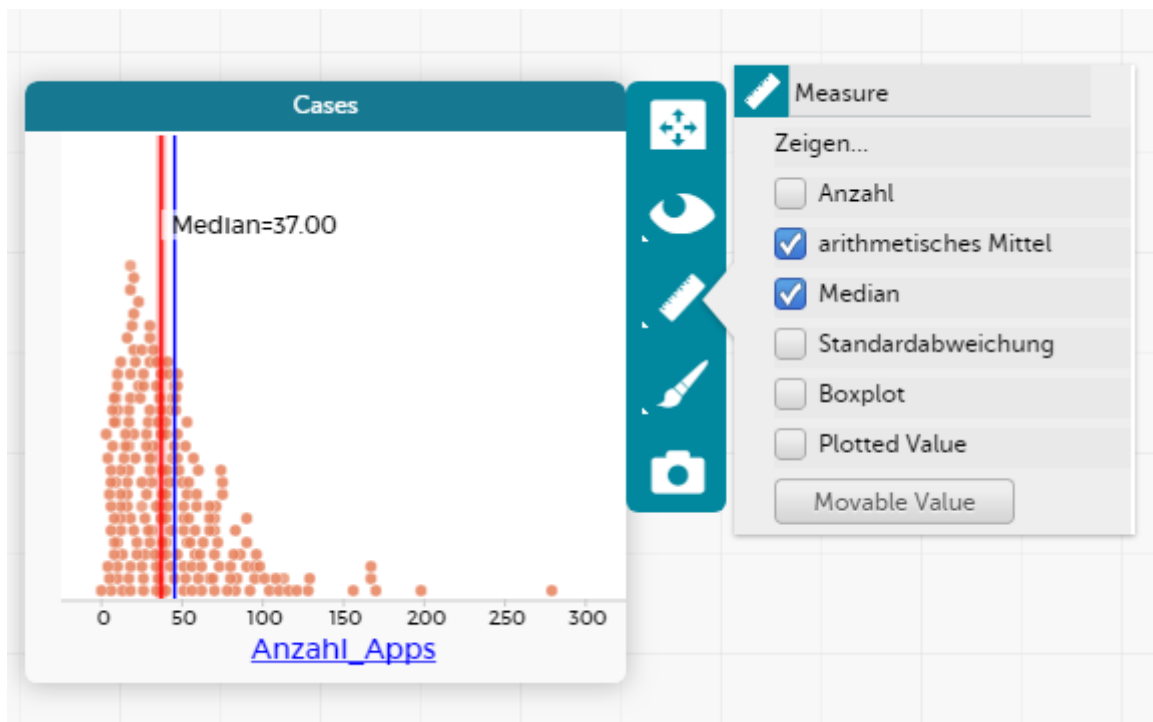
CODAP erstellt dann ein **gestapeltes Punktdiagramm**:



Wir sehen, dass sich die meisten Datenpunkte im Bereich von 0 bis 100 Apps befinden. Einige wenige Datenpunkte liegen deutlich über diese Grenze. Um konkrete Auswertungen vorzunehmen, bietet CODAP weitere Funktionen in der Arbeitsleiste rechts an:

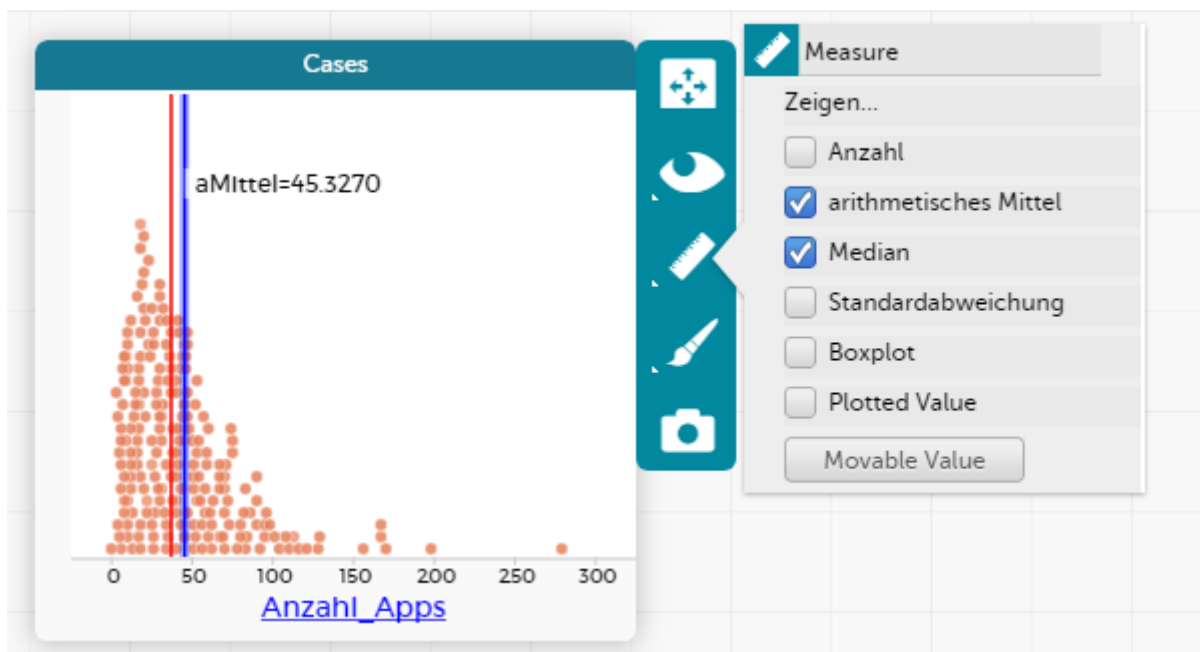


So lassen sich beispielsweise konventionelle Kennzahlen der Verteilung wie der **Median** (in rot) oder das **arithmetische Mittel** (in blau) berechnen. Gehen wir mit der Maus auf eine dieser vertikalen Mittelwertlinien, so können wir die Mittelwerte auch numerisch erfassen:

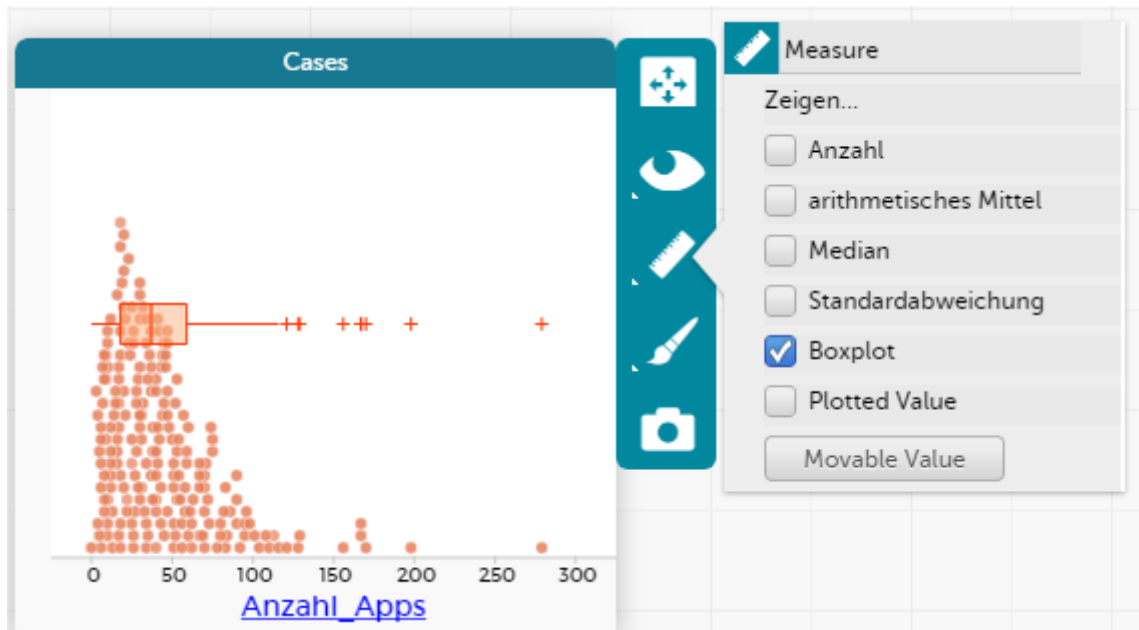


Im **Median** haben die befragten Schülerinnen und Schüler 37 Apps.

Das **arithmetische Mittel** ist größer als der Median (was vor allem an die Ausreißer jenseits der 150 Apps liegen dürfte). Im Durchschnitt haben die befragten Schülerinnen und Schüler 45 Apps, wie wir in der folgenden Darstellung sehen.

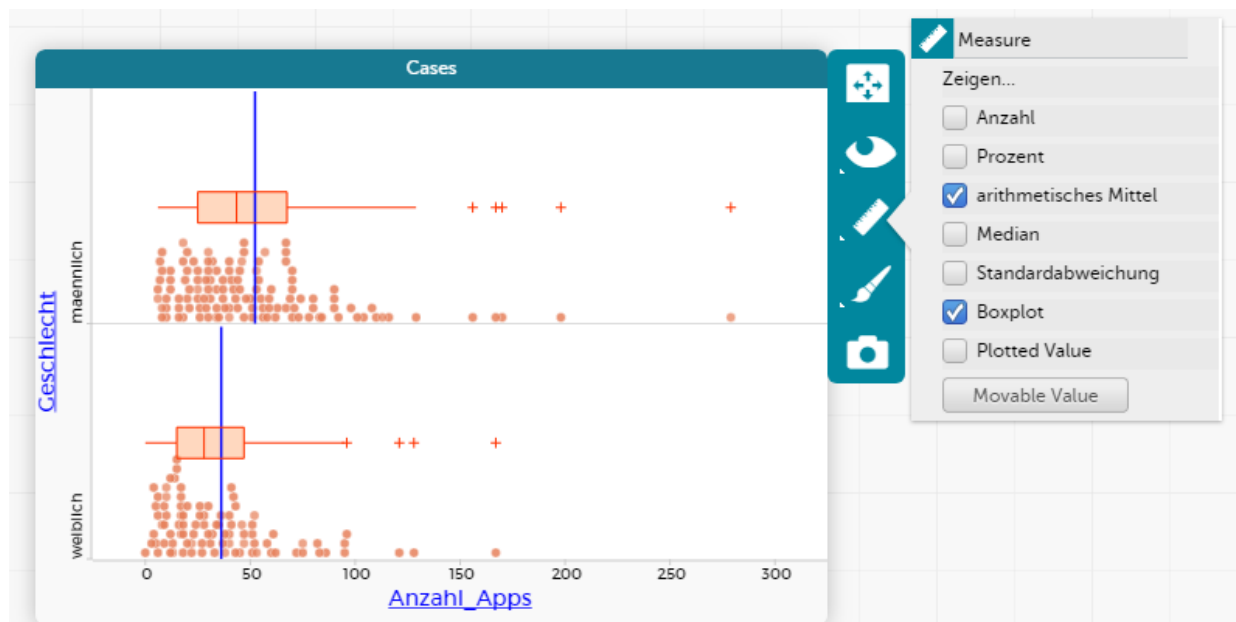


Mithilfe von **Boxplots** können nun weitere Aussagen über die Verteilung des Merkmals **Anzahl_Apps** präzisiert werden.



Beispielsweise lässt sich sagen, dass die mittlere Hälfte der befragten Schülerinnen und Schüler zwischen 18 und 59 Apps auf ihrem Smartphone installiert hat. Mithilfe des Boxplots lassen sich auch gut die Ausreißer (+) erkennen.

Im Weiteren kann sich dann ein **Verteilungsvergleich** beispielsweise mit der Fragestellung „Inwiefern unterscheiden sich die Schülerinnen und Schüler hinsichtlich der Anzahl der Apps auf ihren Smartphones?“ anschließen. Hier kann man dann das Merkmal Geschlecht auf die y-Achse ziehen und erhält die folgende Graphik in CODAP:

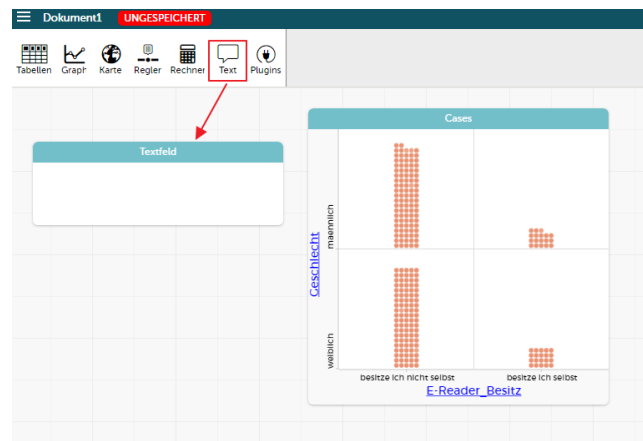


Sowohl beim Betrachten der Boxplots als auch beim Betrachten der arithmetischen Mittelwerte lässt sich sagen, dass in diesem Datensatz die Jungen tendenziell mehr Apps auf ihrem Smartphone haben als die Mädchen. Durchschnittlich haben die Jungen 16 Apps mehr

als die Mädchen. Das spiegelt sich auch beim Vergleich der Mediane (welcher bei dem Vorliegen der Ausreißer in diesem Datensatz, das im Vergleich zum arithmetischen Mittel verlässlichere Vergleichskriterium sein dürfte) wider: Die Jungen haben im Median 15 Apps mehr als die Mädchen. In Bezug auf erste Streuungsunterschiede zwischen beiden Verteilungen lässt sich sagen, dass die Box (der Interquartilsabstand) in der Verteilung des Merkmals „Anzahl_Apps“ bei den Jungen breiter ist als bei den Mädchen und somit die Jungen in diesem Bereich ein heterogeneres Verhalten aufweisen als die Mädchen. Anhand der Verschiebung der Boxen und dem Vergleich der Kennzahlen Q1 und Q3 bestätigt sich der Eindruck, dass in diesem Datensatz die Jungen tendenziell mehr Apps auf ihrem Smartphone haben als die Mädchen.

Text

Um seine Erkenntnisse und Ergebnisse der Datenexploration festzuhalten, bietet es sich an ein Textfeld zu nutzen.



Nach dieser Einführung sollten Sie nun für Ihren zweiten Arbeitsauftrag bereit sein 😊
Diesen finden Sie auf Seite 12.

Ihr zweiter Arbeitsauftrag (10 Minuten)

Untersuchen Sie **eine** der folgenden Fragestellungen im reduzierten JIM-Paderborn-Datensatz in CODAP:

- 1) Inwiefern unterscheiden sich die Jungen und Mädchen hinsichtlich der Instagram-Nutzung? (Merkmale: Geschlecht, Nutzen_Instagram)
- 2) Inwiefern unterscheiden sich die Jungen und die Mädchen im Datensatz hinsichtlich der Anzahl ihrer gespeicherten Kontakte? (Merkmale: Geschlecht, Anzahl_Kontakte)

Gehen Sie bei Ihrer Datenanalyse in CODAP für die Exploration der Fragestellungen 1) und 2) jeweils die folgenden Schritte i, ii und iii durch:

- i. *Erzeugen Sie in CODAP aussagekräftige Graphiken, die Ihnen helfen, Ihre Fragestellung zu beantworten.*
- ii. *Beschreiben (Was ist dargestellt?) und interpretieren (Was bedeutet das für meine Fragestellung?) Sie die Graphiken, die Sie in CODAP erstellt haben. Nutzen Sie zur Dokumentation Ihrer Beschreibung und Ihrer Interpretation das Textfeld in CODAP.*
- iii. *Beantworten Sie jeweils Ihre Fragestellung kurz und prägnant. Dokumentieren Sie Ihre Antworten in einem Textfeld in CODAP.*