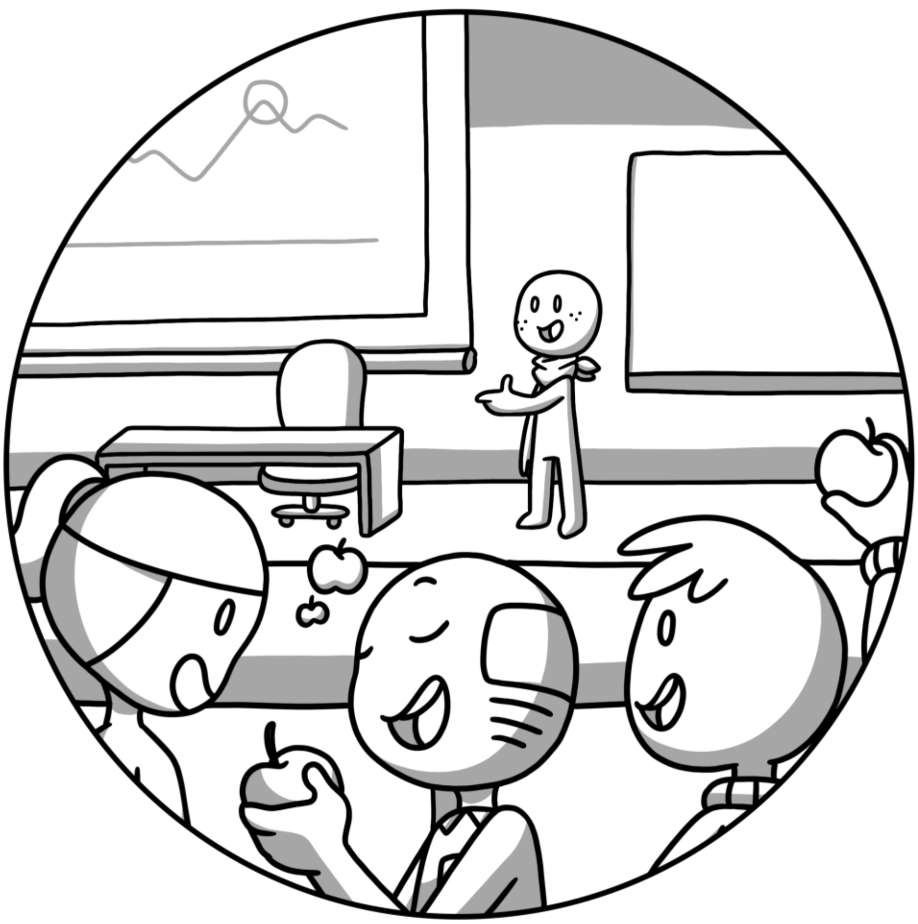
Lagemaße



COMIXPLAIN

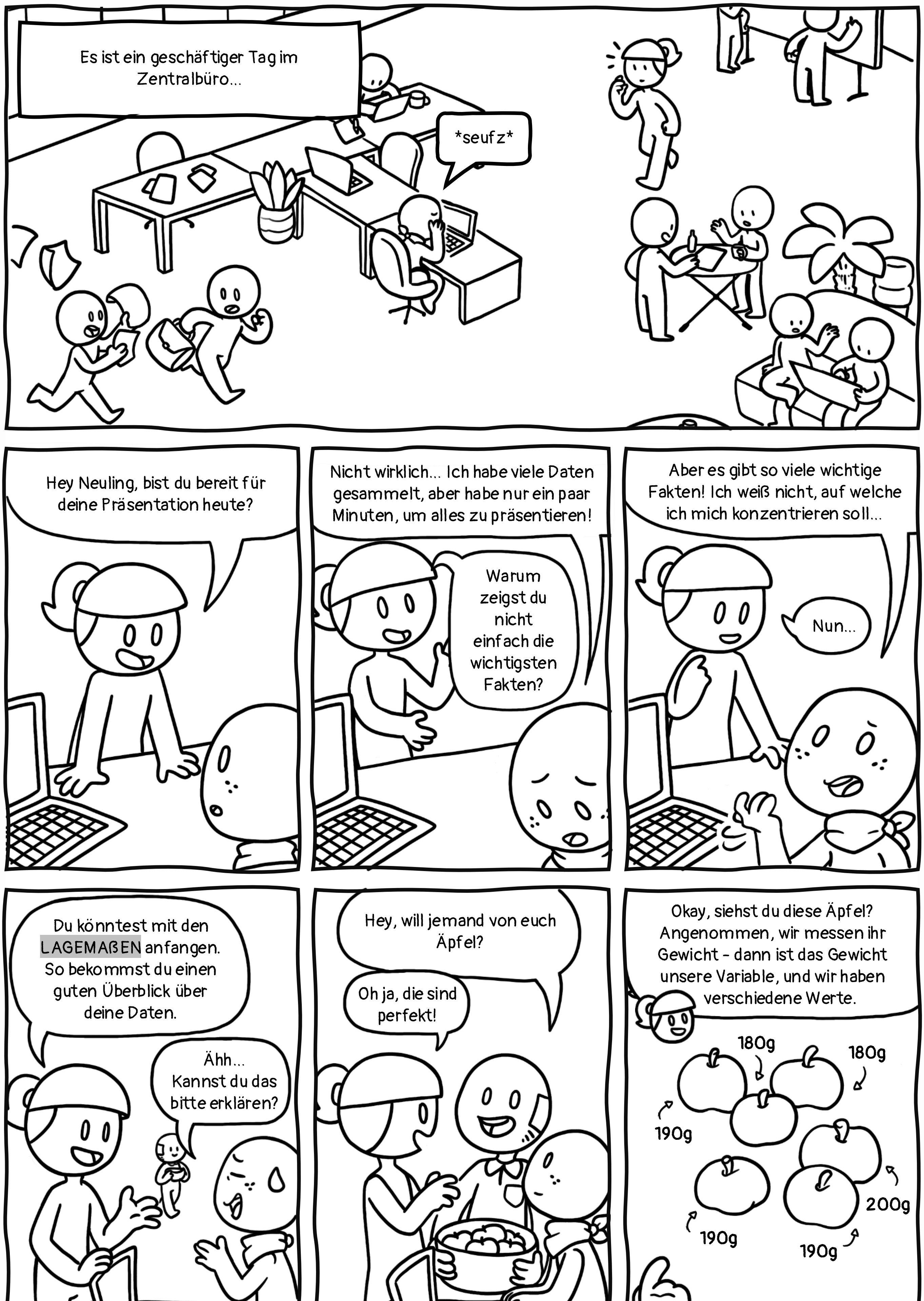
Dieser Comic wurde im Zuge des Forschungsprojekts Comixplain, gefördert von der Fachhochschule St. Pölten im Rahmen des Innovation Call 2022, erstellt.

Projektteam:
Victor-Adriel De-Jesus-Oliveira
Hsiang-Yun Wu
Christina Stoiber
Magdalena Boucher
Alena Ertl

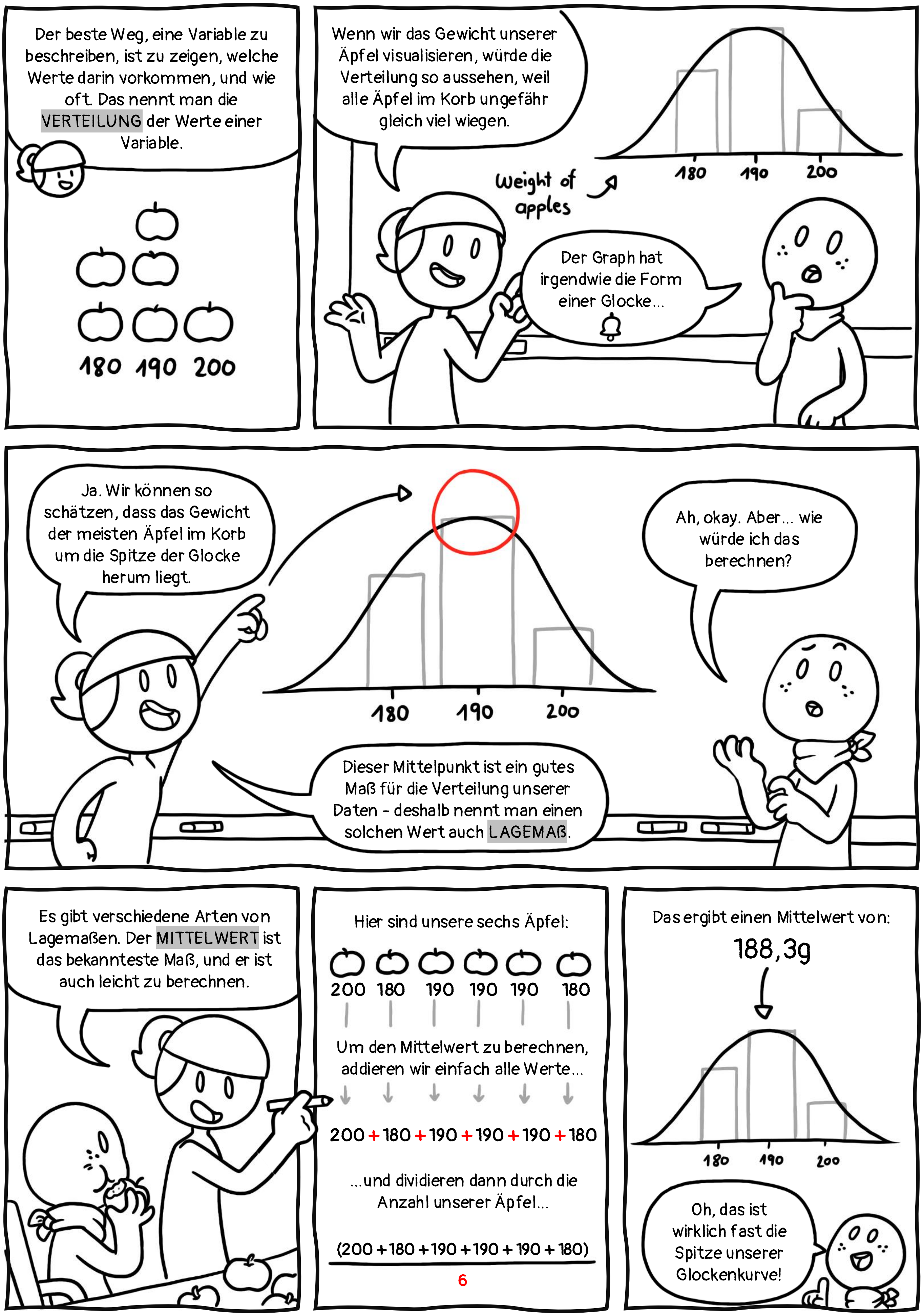
Kontakt:
victor.oliveira@fhstp.ac.at

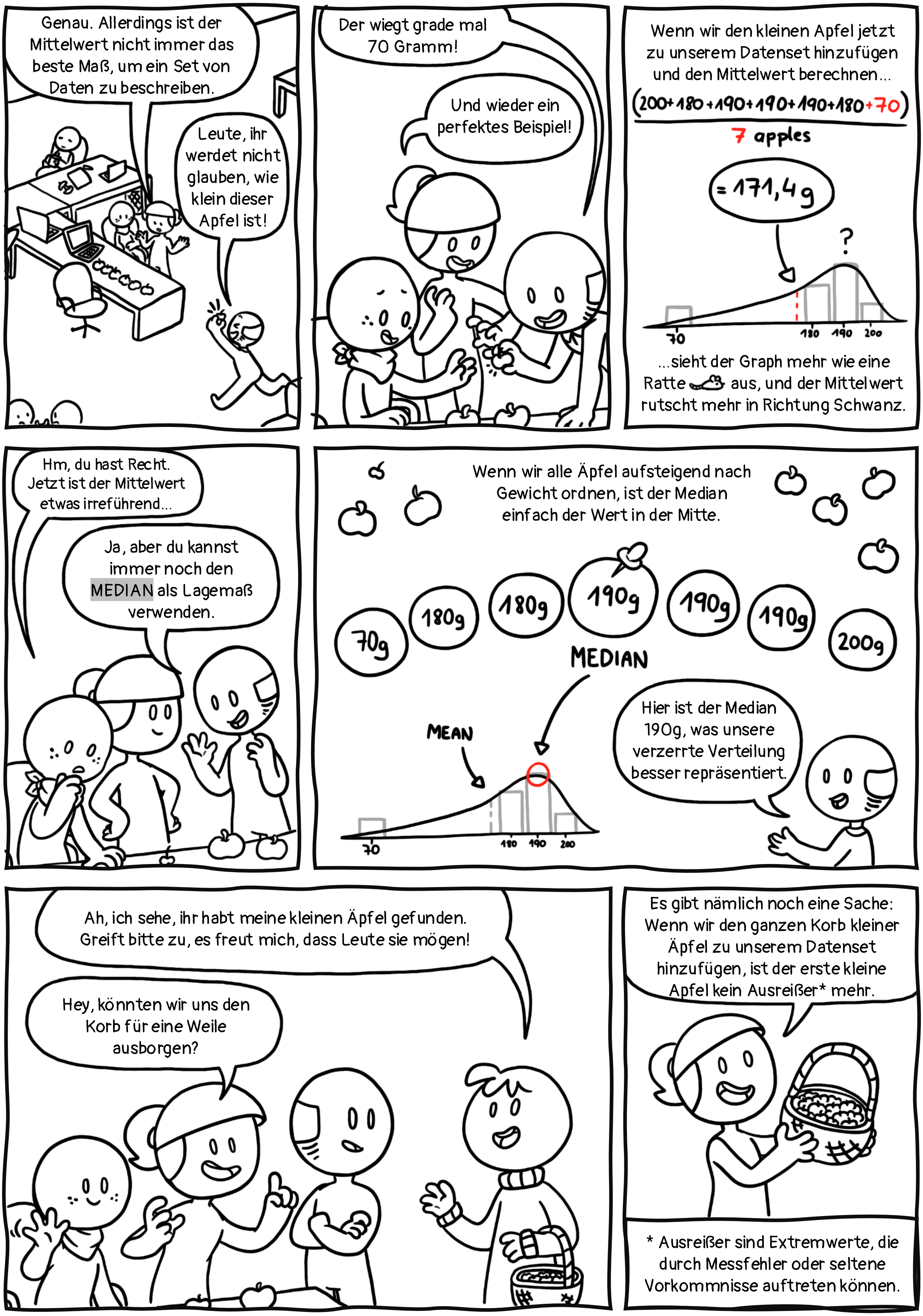
Illustrationen:
Magdalena Boucher & Alena Ertl

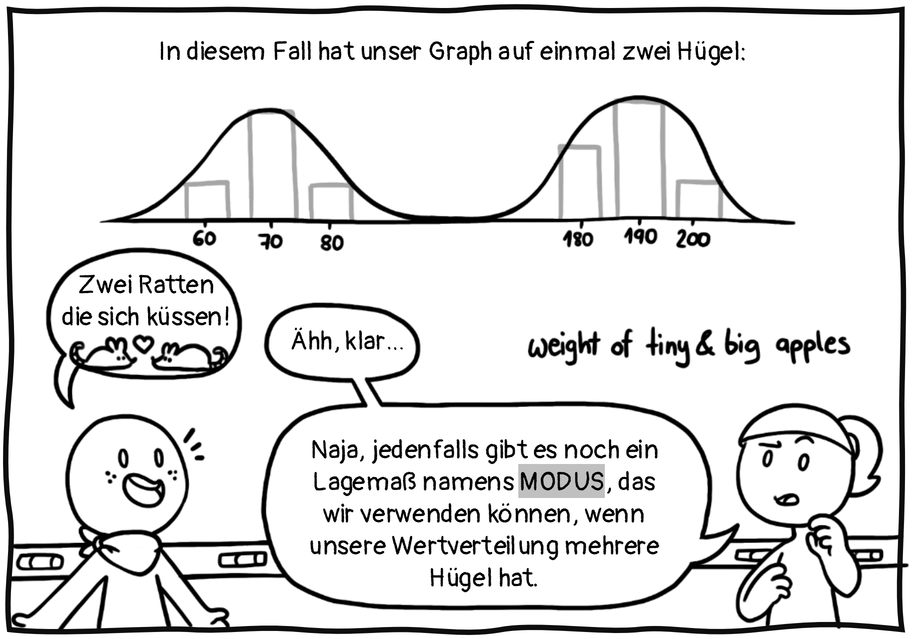
https://fhstp.github.io/comixplain

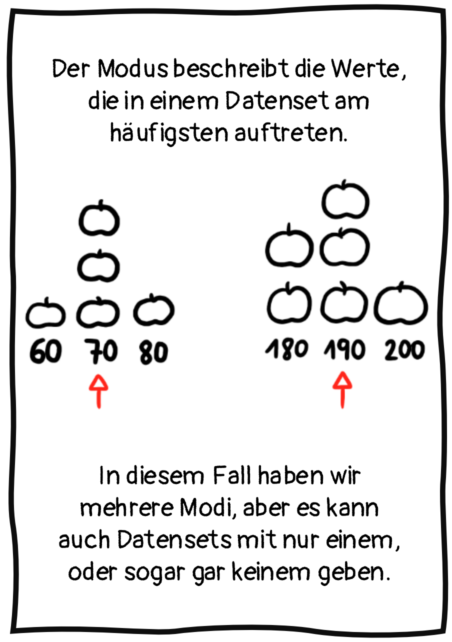


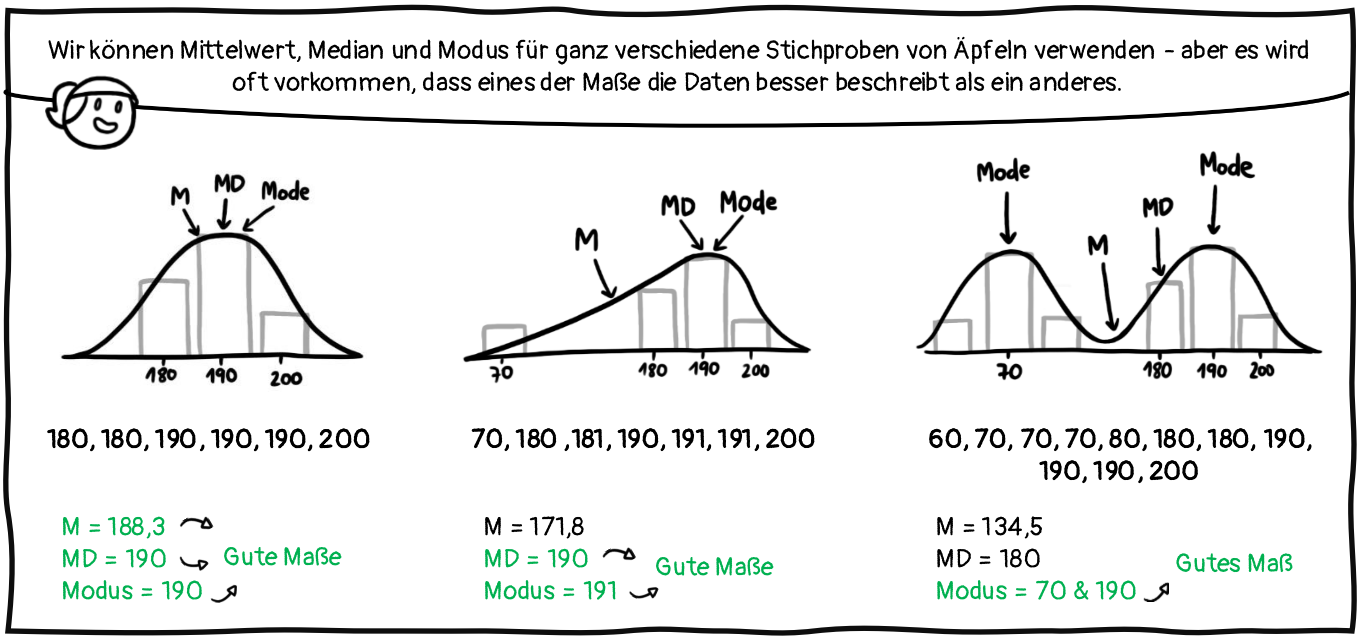
Es ist ein geschäftiger Tag im Zentralbüro...
seufz
Hey Neuling, bist du bereit für deine Präsentation heute?
Nicht wirklich... Ich habe viele Daten gesammelt, aber habe nur ein paar Minuten, um alles zu präsentieren!
Warum zeigst du nicht einfach die wichtigsten Fakten?
Aber es gibt so viele wichtige Fakten! Ich weiß nicht, auf welche ich mich konzentrieren soll...
Nun...
Du könntest mit den LAGEMAßEN anfangen. So bekommst du einen guten Überblick über deine Daten.
Ähh... Kannst du das bitte erklären?
Hey, will jemand von euch Äpfel?
Oh ja, die sind perfekt!
Okay, siehst du diese Äpfel? Angenommen, wir messen ihr Gewicht - dann ist das Gewicht unsere Variable, und wir haben verschiedene Werte.
180g
180g
190g
200g
190g
190g


Der beste Weg, eine Variable zu beschreiben, ist zu zeigen, welche Werte darin vorkommen, und wie oft. Das nennt man die VERTEILUNG der Werte einer Variable.
180 190 200
Wenn wir das Gewicht unserer Äpfel visualisieren, würde die Verteilung so aussehen, weil alle Äpfel im Korb ungefähr gleich viel wiegen.
Weight of apples
180 190 200
Der Graph hat irgendwie die Form einer Glocke...
Ja. Wir können so schätzen, dass das Gewicht der meisten Äpfel im Korb um die Spitze der Glocke herum liegt.
180 190 200
Ah, okay. Aber... wie würde ich das berechnen?
Dieser Mittelpunkt ist ein gutes Maß für die Verteilung unserer Daten – deshalb nennt man einen solchen Wert auch LAGEMAß.
Es gibt verschiedene Arten von Lagemaßen. Der MITTELWERT ist das bekannteste Maß, und er ist auch leicht zu berechnen.
Hier sind unsere sechs Äpfel:
200 180 190 190 190 180
Um den Mittelwert zu berechnen, addieren wir einfach alle Werte...
200 + 180 + 190 + 190 + 190 + 180
...und dividieren dann durch die Anzahl unserer Äpfel...
(200 + 180 + 190 + 190 + 190 + 180) / 6
Das ergibt einen Mittelwert von:
188,3g
180 190 200
Oh, das ist wirklich fast die Spitze unserer Glockenkurve!


Genau. Allerdings ist der Mittelwert nicht immer das beste Maß, um ein Set von Daten zu beschreiben.
Leute, ihr werdet nicht glauben, wie klein dieser Apfel ist!
Der wiegt grade mal 70 Gramm!
Und wieder ein perfektes Beispiel!
Wenn wir den kleinen Apfel jetzt zu unserem Datenset hinzufügen und den Mittelwert berechnen...
(200+180+190+190+190+180+70)
7 apples
= 171,4g
?
70
180
190
200
...sieht der Graph mehr wie eine Ratte aus, und der Mittelwert rutscht mehr in Richtung Schwanz.
Hm, du hast Recht. Jetzt ist der Mittelwert etwas irreführend...
Ja, aber du kannst immer noch den MEDIAN als Lagemaß verwenden.
Wenn wir alle Äpfel aufsteigend nach Gewicht ordnen, ist der Median einfach der Wert in der Mitte.
70g
180g
180g
190g
190g
190g
200g
MEDIAN
MEAN
70
180
190
200
Hier ist der Median 190g, was unsere verzerrte Verteilung besser repräsentiert.
Ah, ich sehe, ihr habt meine kleinen Äpfel gefunden. Greift bitte zu, es freut mich, dass Leute sie mögen!
Hey, könnten wir uns den Korb für eine Weile ausborgen?
Es gibt nämlich noch eine Sache: Wenn wir den ganzen Korb kleiner Äpfel zu unserem Datenset hinzufügen, ist der erste kleine Apfel kein Ausreißer* mehr.
* Ausreißer sind Extremwerte, die durch Messfehler oder seltene Vorkommnisse auftreten können.







Nutzer ID	Herz-Frequenz (bpm)	Nutzungszeit	Bewertung
1	45	13:00	1
2	50	9:00	5
3	55	10:00	3
4	57	9:00	4
5	63	14:00	5
6	70	15:00	5
7	65	16:00	4
8	75	15:00	2

Bevor du umblätterst: Versuche, den Mittelwert, Median und Modus für jede Variable in der Tabelle zu berechnen. Entscheide, welches Maß am besten für welche Variable geeignet ist. Du kannst auf dieser Seite Notizen machen!

Du kannst die Lösungen auf der nächsten Seite zum Vergleich anschauen und auf dieser Seite weitere Notizen machen.

HERZFREQUENZ

Berechnung des MITTELWERTS:

$$\frac{45+50+55+57+63+70+65+75}{8 \text{ users}} = \frac{480}{8} = 60 \text{ bpm}$$

Berechnung des MEDIANS:

45 50 55 [57 63] 65 70 75

Wenn es zwei mittlere Werte gibt, dann berechnet sich der Median aus dem Mittelwert der beiden Werte:
(57+63)/2 = 60bpm)

Berechnung des MODUS:

45, 50, 55, 57, 63, 70, 65, 75

Jeder Wert kommt nur ein Mal vor - es gibt also keinen Modus!

HÄUFIGSTE NUTZUNGSZEIT

Berechnung des MITTELWERTS:

$$\frac{9+9+10+13+14+15+15+16}{8 \text{ Nutzer*innen}} = \frac{101}{8} = 12{,}6?$$

Nutzungszeit ist kein quantitativer Wert - daher macht die Berechnung von Mittelwert und Median keinen Sinn!

Berechnung des MEDIANS:

9 9 10 [13 14] 15 15 16 = 13,5?

Berechnung des MODUS:

9:00	10:00	13:00	14:00	15:00	16:00
2x	1x	1x	1x	2x	1x

= 2 Modi: 9:00 & 15:00

STERNEBEWERTUNG

Berechnung des MITTELWERTS:

$$\frac{1+2+3+4+4+5+5+5}{8 \text{ users}} = \frac{29}{8} = 3{,}6 \text{ stars}$$

Berechnung des MEDIANS:

1 2 3 [4 4] 5 5 5 = 4 stars

Berechnung des MODUS:

1	2	3	4	5
1x	1x	1x	2x	3x

= 5 stars

Quellen:

Downey, A. (2014). Think stats: exploratory data analysis. O'Reilly Media, Inc.

Field, A. (2022). An adventure in statistics: The reality enigma. Sage.