

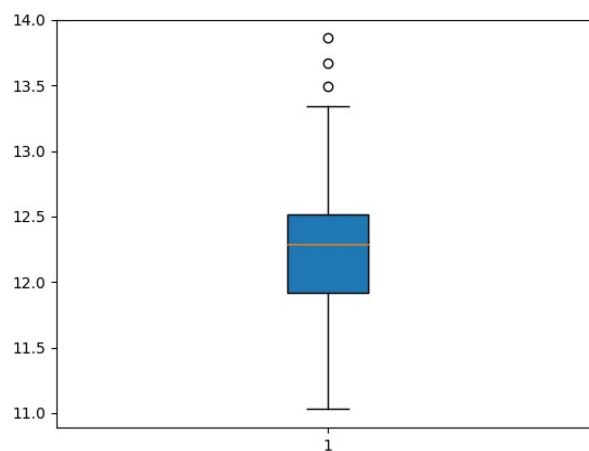
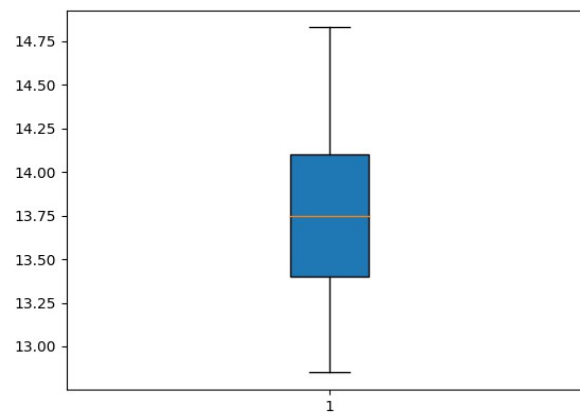
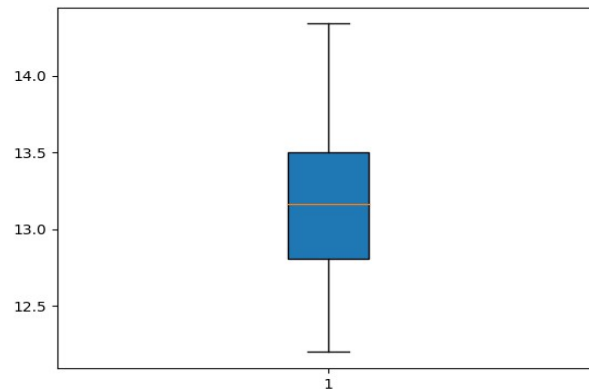
A1)

Boxplots eignen sich zum Vergleich von Bereich und Verteilung von Gruppen mit Zahlendaten.

Vorteile: Boxplots strukturieren große Datenmengen und machen Ausreißerwerte sichtbar.

Man kann anhand eines Boxplots wichtige Lageparameter erkennen.

Zur ersten Sichtung der Daten ist dies hilfreich, daher sollte dieser ins Skript



2)

Blocking tracking ads?	<u>No</u>
Blocking invisible trackers?	<u>No</u>
Protecting you from <u>fingerprinting</u> ?	<u>Your browser has a unique fingerprint</u>

Sind Sie einzigartig unter den Nutzer\*innen dieser Webseite :

Nein.

Haben Sie einen unique fingerprint?:

Ja.

Ich habe einen unigen Fingerprint. Was ich unter diesem Kontext verstehe ist, dass jedesmal wenn mein Computer Anfragen an den Web schickt, werden Informationen weitergeleitet in Form von Header mit jeweiligen Attributen wie welchen Browser/Sprache ich nutze.

Was bedeutet es, wenn dort so etwas wie "Bits of identifying information: 1.13 One in x browsers have this value: 2.19" steht?:

Ich habe 17.73 Bits pro Zeichen

In welchem Zusammenhang steht dieser Begriff "Bits of information" mit der Entropie, die wir kennen gelernt haben?:

Die Entropie gibt den Informationsgehalt wieder, je höher die Dichte desto schwieriger lassen sich verlorene Zeichen wiederherstellen, d.h wir senden pro Zeichen xx Bits, dementsprechend ist auch unsere Auftrittswahrscheinlichkeit  $p_i$  sehr klein. Denn wenn man eine hohe Entropie hat, dann gilt auch, dass  $p_i$  sehr klein ist. Den Informationsgehalt wie in meinem Falle 17.73 gibt somit eindeutig, dass die Auftrittswahrscheinlichkeit eines Nutzers genau denselben Zustand wie ich hat, sehr niedrig ist.

Was ist hier die Zufallsvariable?:

Die Zufallsvariable wäre nach meinem Bauchgefühl eher ein Zufallsvektor mit jeweils verschiedenen univarianten Einträgen  $\rightarrow X = (\text{Browser, Sprache, Ort, ...})$

Was lernen Sie (oder ein\*e fiktive\*r Studierende\*r, für die das neu ist) daraus für die Erfassung und Verarbeitung von angeblich anonymisierten personenbezogenen Daten?:

Man lernt daraus, dass im Internet nichts wirklich anonym ist und im Internet nie wirklich alles 'verloren geht'.

3)

1: weight-height.csv

2: <https://www.kaggle.com/datasets/mustafaali96/weight-height>

3: (10001, 3)

4: Sind Männer größer als Frauen?, Sind Männer schwerer als Frauen?, Gibt es einen linearen Zusammenhang zwischen Größe und Gewicht?, Größe und Gewicht sind normalverteilt ?

5:

**minHeight**: 54.2631333250971

**maxHeight**: 78.9987423463896

**meanHeight**: 66.36755909127837

**minWeight**: 64.700126712753

**maxWeight**: 269.989698505106

**meanWeight**: 161.44035521875003

6: Meine Erwartung zu der Größe wurde erfüllt, mein Bauchgefühl sagte, die Größe ist eindeutig normalverteilt und in der Tat, ist sie es.

Am Scatter plot erkenne ich, dass eine positive Korrelation zwischen der Größe und Gewicht gibt, meine Vermutung stimmte ebenso.

Was mich überrascht hat ist, dass das Gewicht nicht eindeutig normalverteilt ist sondern bimodal bzw. eine andere Form der Normalverteilung.

7: Mich hätte es interessiert, ob diese Korrelation auch wirklich kausal ist oder nur wegen der Stichprobe dieser Zusammenhang existiert!

Ich hätte gerne das Verhältnis von Männern und Frauen näher betrachtet, aber leider sind die Variablen spalten alle auf NaN gesetzt. Dies wird mir dadurch erschwert. Mithilfe Pandas müsste aber dies möglich sein.

Hierzu 3 Plots:

