

SYT7 4xHIT 23/24 / Vertiefung Data Science / [GK] DaSci 7.2 Explorative Datenanalyse und Visualisierung – 4h

# [GK] DaSci 7.2 Explorative Datenanalyse und Visualisierung – 4h

✓ Done: View ✓ Done: Make a submission ✓ Done: Receive a grade

Opened: Monday, 18 September 2023, 12:00 AM Due: Wednesday, 20 December 2023, 12:00 AM

# Data Science "Explorative Datenanalyse und Visualisierung" - Taskdescription

nheiten sollen als Einführung in die explorative Datenanalyse und Visualisierung dienen. Zur Einführung in R, RStudio oder Python ist der Link aus den Theorie Einheiten empfohlen.

#### Ziele

Das Ziel ist es das Grundwerkzeug zur Datenexploration und Visualisierung zu verstehen, einsetzen und interpretieren zu können.

#### Voraussetzungen

- . Kenntnis von Markdown, LaTeX und Grundkenntnisse in R und Python
- Kunktionstüchtige Installation von R, RStudio, Python auf eurem Rechner, virtueller Maschine etc.
   Kenntnis von Variablentypen und Schätzern bzw. Methoden zur Informationszusammenfassung für unterschiedliche Variablentypen (Häufigkeiten, Lage, Streuung, etc.)
   Kenntnis von Möglichkeiten zur Visualisierung unterschiedlicher Datentypen

#### Aufgabe 1 (GK)

Exploration numerischer Daten (Wirtschaftsdaten der US Bundesstaaten)

- 2. Ermittle mithilfe geeigneter Schätzer für die Lage (arithmetischer Mitte vert und Median sollen verglichen werden) und Streuung (Standardabweichung und Interquartilsdistanz sollen verglichen werden) der ersten 5 Variablen: Population, Income, Illiteracy, Life Exp(ectancy) und Murder.
- 3. Stelle die Daten der ersten 5 Variabhen, Population, Income, Illiteracy, Life Exp(ectancy) und Murder in geeigneter Weise graphisch dar, indem du Boxplot, Histogramm mit Dichteschätzung, ECDF und QQ-Plot verwendest.

  4. Begründe anhand der graphischen Darstellungen, ob es sich symmetrische oder schiefe Datenverteilungen handelt. Begründe anhand der graphischen Darstellungen, ob schwere oder leichte Ränder vorliegen (und auf welch

  5. Bestimme anhand graphischen Darstellungen aus Punkt 3. und der Erkenntnisse aus 4., ob Ausreißer vorliegen und welche Punkte dies sind.

## Anfgabe 2 (EK)

Exploration kategorialer Daten (Überlebende der Titanic)

- 1. Lade den Datensatz 'Titanic', welche im R base Paket inkludiert ist. Beschreibe die Daten anhand der internen Hilfe. (empfohlen sind dafür R Befehle: ?, help() )
- 2. Untersuche den "Titanic" Datensatz mittille von Tabellen von absoluten und relativen Häufigkeiten der Ränder herundragebrochen auf je 2 der 3 Dimensionen von Variablen: Stelle dabei die Überlebendenden aufgeteilt nach dem Geschlecht, nach dem Alter und nach der Klasse als zweidimensionale Tabelle dar. Gib die Aufteilung der Überlebenden nach Geschlecht und Klasse zugleich, nach Alter und Klasse zugleich wieder. (empfohlen sind dafür die R Befehle: table(), flable(), colSums, rowSums(), apply(), xtable())

  3. Stelle die Daten in geeigneter Weise mithille von barplots und mosaicplots graphisch dar. (empfohlen sind dafür R Befehle: barplot(), mosaicplot(), assocplot())
- 4. Interpretiere die Tabellen und ihre dazugehörenden Grafiken. Begründe anhand der quantitativen Ergebnisse, ob Frauen und Kinder zuerst bei der Seenotrettung berücksichtigt wurde. 5. Erkläre, wo sich Anzeichen für Simpson's Paradoxon finden lassen.

#### **Abgabe**

Das Protokoll ist als PDF-Dokument abzugeben vorzulegen, welches die graphischen Darstellungen und Interpretationen in ganzen deutschen Sätzen enthält

## Bewertung

Gruppengrösse: 1-2 Pers

## Anforderungen überwiegend erfüllt

- \* aktuelle Markdown- oder LaTeX-Potokolivoriage aus Github bzw. Moodle verwendet
   \* grundlegende Beschreibung und Verwendung der im Unterricht angeführten Begriffe: ordinale, nominal, metrisch rational, metrisch intervallskallert; absolute und relative Häufigkeiten; Mittelwert, Median, Varianz, Standardabweichung, Interquartilsdistanz; Symmetrie, Schiefe, schwa Ränder, Ausreißer; Histogramm, Boxplot, QQ-Plot
- · Codebeispiele referenziert
- Dokumentation aller Inhalte des DataCamp Kurses

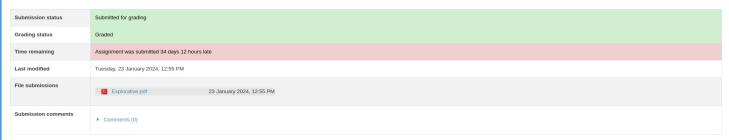
## Anforderungen zur Gänze erfüllt

- Verbale Beschreibung und Erklärung aller angeführter Begriffe und deren Anwendung in konkreten Beispielen in vollständigen deutschen Sätzen
   ausführliche Codebeispiele und Visualisierungen dokumentiert

## Quellen

Edit submission Remove submission

### Submission status



## Feedback



◀ [EK] Shiny Web APP

Jump to.

[EK] - Visualisierung ▶