

Wichtiger Hinweis: Sie dürfen den Datensatz ausschließlich im Rahmen dieser Lehrveranstaltung verwenden. Es ist untersagt, den Datensatz oder Informationen über ihn weiterzuverbreiten. Bitte löschen Sie den Datensatz nach den Präsentationen am 28.02.2024 bzw. nach Abgabe des Berichts am 18.03.2024.

Im Moodle finden Sie den Datensatz `kuehe.Rdata`. Dieser enthält einen vorverarbeiteten und gekürzten Auszug des Originaldatensatzes aus dem Projekt optiKuh, welches vom 01.09.2014 bis 30.06.2018 lief. Weitere Informationen zum optiKuh-Projekt finden Sie auf der Homepage <https://www.optikuh.de/>. Im Folgeprojekt optiKuh2 (Laufzeit: 01.09.2019 – 30.11.2021) war die Zielsetzung, diese Daten zu verwenden, um einen Algorithmus zur Vorhersage des peripartalen Erkrankungsrisikos bei Milchkühen zu entwickeln. Der gegebene Datensatz enthält 1 228 Beobachtungen von Milchkühen zu 34 Merkmale, diese umfassen im Einzelnen:

- **farm:** Anonymisierte Betriebsnummer
- **lom:** Ohrenmarkennummer der Kuh
- **gebdat:** Geburtsdatum der Kuh
- **rasse:** Rasse der Kuh
- **datum:** Datum der Beobachtung
- **lnr:** Laktationsnummer: Als Laktation wird der Zeitraum bezeichnet, indem eine Kuh Milch gibt. Dieser liegt zwischen der Kalbung, also der Geburt eines Kalbs, und dem Tockenstellen, d. h. der Beendigung der Laktation. Die Laktationsnummer gibt nun an in der wievielten Laktation sich eine Kuh befindet.
- **kaldat:** Letztes Kalbedatum: Wann hat die Kuh ihr letztes Kalb geboren?
- **ltg:** Laktationstag: Gibt den Tag innerhalb einer Laktation an.
- **gew:** Letztes gemessenes Körpergewicht der Kuh in kg
- **bcs:** Letzter gemessener Body-Condition-Score: Ist ein Maß zur Körper-Konditionsbeurteilung. Er dient zur Einschätzung der Energiereserven einer Kuh. Ein niedriger Wert bedeutet, dass die Kuh (zu) mager ist und ein hoher Wert, dass sie (zu) fett ist.
- **rfd:** Letzte gemessene Rückenfettdicke in mm
- **mkg:** Mittlere Milchmenge in kg aus der vorherigen Laktation.
- **bhb:** Gemittelte β -Hydroxybutyrat-Werte (BHB) in mmol/L aus der vorherigen Laktation.
- **glukose:** Gemittelte Glukosewerte in mmol/L aus der vorherigen Laktation.
- **nefa:** Gemittelte nicht veresterte freie Fettsäure (NEFA) in μ mol/L aus der vorherigen Laktation.
- **insulin:** Gemittelte Insulinwerte in μ U/mL aus der vorherigen Laktation.
- **igf1:** Gemittelte Werte des Insulin-ähnlicher Wachstumsfaktor-1 (IGF-1) in ng/mL aus der vorherigen Laktation.
- **ca:** Gemittelte Kalziumwerte in mmol/L aus der vorherigen Laktation.
- **abgdat:** Abgangsdatum: Wann hat die Kuh den Betrieb verlassen?
- **abgrund:** Abgangsgrund: Warum hat die Kuh den Betrieb verlassen?
- **mastitis:** Ist in der vorliegenden Laktation eine Mastitis aufgetreten (0 = nein, 1 = ja)?
- **mastitis_tage:** Nach wie vielen Tag ist eine Mastitis aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.
- **metritis:** Ist in der vorliegenden Laktation eine Metritis aufgetreten (0 = nein, 1 = ja)?
- **metritis_tage:** Nach wie vielen Tag ist eine Metritis aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.

- **hypocal**: Ist in der vorliegenden Laktation eine Hypocalcämie aufgetreten (0 = nein, 1 = ja)?
- **hypocal_tage**: Nach wie vielen Tag ist eine Hypocalcämie aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.
- **ketose**: Ist in der vorliegenden Laktation eine Ketose aufgetreten (0 = nein, 1 = ja)?
- **ketose_tage**: Nach wie vielen Tag ist eine Ketose aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.
- **nachgebver**: Ist in der vorliegenden Laktation eine Nachgeburtsverhaltung aufgetreten (0 = nein, 1 = ja)?
- **nachgebver_tage**: Nach wie vielen Tag ist eine Nachgeburtsverhaltung aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.
- **zyklus**: Ist in der vorliegenden Laktation eine Zyklusstörung aufgetreten (0 = nein, 1 = ja)?
- **zyklus_tage**: Nach wie vielen Tag ist eine Zyklusstörung aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.
- **labmagen**: Ist in der vorliegenden Laktation eine Labmagenverlagerung aufgetreten (0 = nein, 1 = ja)?
- **labmagen_tage**: Nach wie vielen Tag ist eine Labmagenverlagerung aufgetreten? Falls keine Mastitis innerhalb einer Laktation aufgetreten ist, so ist der Wert gleich der maximalen Beobachtungszeit.

Ihre Aufgabe ist es, den Datensatz sinnvoll graphisch zu untersuchen und darzustellen. Bereiten Sie dazu in Gruppenarbeit eine Präsentation vor. Gruppen mit drei Personen haben dabei 15 Minuten und Vierergruppen 20 Minuten Zeit. Zudem sollte jedes Gruppenmitglied einen Teil der Arbeit vorstellen.

Wenn Sie **Wissenschaftsjournalismus** oder **Data Science** studieren oder sich im **Erasmus** Studiengang befinden, schreiben Sie *zusätzlich* in Einzelarbeit einen Bericht von 8 – 9 Seiten Länge.

Suchen Sie sich **einige wenige** Aspekte bzw. Fragestellungen aus. Diese sollten Sie sinnvoll betrachten und darstellen.

Versuchen Sie nicht, jede Variable und jeden denkbaren Zusammenhang in seiner Gänze zu beleuchten. Das wird Ihnen nicht gelingen und ist auch nicht das Ziel dieses Abschlussprojektes. In diesem Sinne:

- Überlegen Sie sich *vor* der Visualisierung, was genau Sie untersuchen möchten und formulieren Sie Untersuchungsziele bzw. Fragestellungen. Gehen Sie also zielgerichtet vor.
- Versuchen Sie nicht, jede kennengelernte Darstellungsform zu verwenden. Zum einen ist das zu viel und zum anderen sollte die gewählte Darstellungsmethode zu Daten und Untersuchungsziel passen.

Es gilt: Erläutern Sie, wie Sie vorgegangen sind. Welche Probleme gab es auf dem Weg zur finalen Graphik? Welche Vor- und Nachteile gibt es mit der von Ihnen gewählten Darstellungsform? Begründen Sie Ihr Vorgehen. Warum wählen Sie welche Darstellungsmethode, Graphikeinstellung, Farbe etc.? Interpretieren Sie Ihre Graphiken zudem sinnvoll. Wie lässt sich Ihre Fragestellung bzw. Ihr Untersuchungsziel mithilfe der Graphik beantworten?

Insgesamt sind Sie in der Bearbeitung der Aufgabe sehr frei:

- Sie dürfen sich den R-Paketen Ihrer Wahl bedienen, diese sollten dann auch *kurz* vorgestellt und kritisch hinterfragt werden (z. B. in Bezug auf Standardeinstellungen und Limitationen).

- Wenn Sie eine trickreiche Plotfunktion schreiben, dürfen Sie diese gerne vorstellen und Ihr Vorgehen erläutern. Gleiches gilt für besonders schönen oder effizienten Code.
- Die in der Vorlesung und Übung behandelten Darstellungsformen stellen natürlich nur einen Teil aller existierenden Darstellungsformen dar. Selbstverständlich sind Sie frei, bei der Untersuchung des Datensatzes auch nicht erwähnte Methoden einzusetzen, wenn sich daraus ein Mehrwert ergibt. Stellen Sie dann die gewählte Darstellungsmethode kurz vor und erläutern Sie wie diese konstruiert sowie in R umgesetzt werden kann.

Hinweise zur Präsentation:

- Stellen Sie zu Beginn kurz Ihre Zielsetzung und die betrachteten Variablen vor.
- Berücksichtigen Sie die vorgestellten Konzepte zur Erstellung von Graphiken, d.h. Graphiken sollten sinnvoll und sauber beschriftet sein bzw. Farben sollten mit Bedacht gewählt werden.
- Genutzte Software inklusive Pakete sollte genannt werden. Sie können auf den Eintrag, der im Literaturverzeichnis verwendet werden sollte, mithilfe von `citation('Paketname')` zugreifen.

Hinweise zum Bericht:

- Der Bericht sollte sich an der Präsentation orientieren. Sie müssen jedoch nicht jeden Aspekt der Präsentation auch in Ihrem Bericht aufgreifen. Versuchen Sie in jedem Fall Hinweise aus der Diskussion, die im Anschluss an die Präsentation stattfindet, umzusetzen.
- Finden Sie eine gute Balance zwischen Text und Graphiken. Geben Sie bitte nicht seitenweise Graphiken ab, die nicht erläutert werden. Auf jede Graphik, die dargestellt wird, sollte im Text Bezug genommen werden.
- Hinweise zum Schreiben eines Berichts finden Sie in Moodle unter *Leitfaden_Berichterstellung.pdf*. Orientieren Sie sich bei der Erstellung Ihres Berichts an diesem Leitfaden.
- Zusätzlich können Sie die Berichtsvorlage unter *Berichtsvorlage.tex* verwenden.

Die Präsentation finden am **Dienstag, den 27.02.2024 ab 14:00 Uhr** und am **Mittwoch, den 28.02.2024 ab 14:00 Uhr im Raum M/E21** statt. Die einzelnen Vortragstermine werden separat bekanntgegeben. Seien Sie bitte **nur** an dem Tag Ihres eigenen Präsentationsvortrags anwesend.

Laden Sie Ihre Präsentationsfolien als **pdf**-Datei mit zugehörigem (lauffähigen) R-Code bis spätestens **12:00 Uhr am 27.02.2024** im Moodle-Raum hoch, unabhängig davon an welchem Tag Sie vortragen.

Die Berichtabgabe (inklusive lauffähigem R-Code) ist bis **Montag, dem 18.03.2023 um 18:00 Uhr im Moodle-Raum** möglich.