

Aufgabe 9 - Datenqualität (20 Punkte)

1. Laden Sie die Salzburger Stationsdaten MON Datensatz_176703_202312.csv in ein R-Projekt

FILTER

107/1011

Bundesland

KARTE

		Anzahl	Klima-ID ↑	Name	Startdatum
▼	<input type="checkbox"/>	4?	7	BAD GASTEIN	Jan. 1935
▼	<input type="checkbox"/>	3?	14	BISCHOFSHOFEN	Jan. 1936
▼	<input type="checkbox"/>	3?	47	KRIMML	Jan. 1891
▼	<input type="checkbox"/>	4?	60	MATTSEE	Juli 1949
▼	<input type="checkbox"/>	3?	68	OBERTAUERN	Juni 1971
▼	<input type="checkbox"/>	6?	73	RADSTADT-TANDALIER	Jan. 1936

Zeilen pro Seite

6

1 - 6 von 107

<< < > >>

+ -

Suche

Startdatum

Enddatum

Source: [Geosphere Data Hub](#)

Aufgabe 9 - Datenqualität (20 Punkte)

2. Sehen Sie sich die Struktur der Daten genau an. Kommentieren Sie kurz, welche Attribute im Datensatz enthalten sind. Welche zeitliche Auflösung weist der Datensatz auf? (2P)
3. Entfernen Sie Zeilen ohne Erfassung (`tidyr::drop_na()`)
4. Begrenzen Sie die Zeitreihe auf max. 2 Jahre.
5. Berechnen Sie aus den Daten vier neue Variablen:
 - Den arithmetischen Mittelwert der Temperatur (t = Lufttemperatur 2m über Boden in °C) aus allen verfügbaren Stationsmessungen für jede Zeiteinheit - `group_by(time)`
 - Die Stichprobengröße als Anzahl der verfügbaren Stationsmessungen (n)
 - Die obere 95% KI-Grenze
 - Die untere 95% KI-Grenze

Aufgabe 9 - Datenqualität (20 Punkte)

6. Stellen Sie alle vier berechneten Variablen in einem einfachen Liniendiagramm mit der Zeit auf der x-Achse in R dar. Syntax: (10P für 3-6)

```
ggplot2::ggplot(data = ., aes(x = <time>)) +  
  ggplot2::geom_line(aes(y = <mean temperature>)) +  
  ggplot2::geom_line(aes(y = <KI upper bound>)) +  
  ggplot2::geom_line(aes(y = <KI lower bound>)) +  
  ggplot2::geom_line(aes(y = <sample size>))
```

Aufgabe 9 - Datenqualität (20 Punkte)

7. Interpretieren Sie die Bedeutung des Konfidenzintervalls: Was bedeutet ein schmales/breites Konfidenzintervall? Wird das Konfidenzintervall schmaler oder breiter, wenn das Konfidenzniveau von 95% auf 99% erhöht wird? In welchem Zusammenhang steht das Konfidenzintervall mit Standardabweichung der Stichprobe bzw. mit der Größe der Stichprobe? (4P)
8. Wie könnten wir vorgehen, um die Konfidenz in unsere Messung der mittleren Lufttemperatur im Bundesland Salzburg zu erhöhen? Es könnten zusätzliche Messstationen aufgestellt werden, allerdings kostet das auch Geld. Wie könnten bei begrenzten Geldmitteln die wenigen Messstationen im Bundesland verteilt werden, um der räumlichen Variabilität der Messgrößen möglichst gut gerecht zu werden? (4P)

Dokumentieren Sie die Schritte bzw. die Antworten auf die Fragen inkl. R Plots in einem PDF. Reichen Sie das PDF inklusive R Skript als ZIP-File mit Ihrem Namen versehen unter Aufgabe 9 in Blackboard bis zum **25.12.2025** ein.