



# Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2024/25

Belinda Fleischmann

Herzlich willkommen!



Aufnahme läuft.

## (1) Einführung

Formalia

Grundbegriffe der Informatik

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

## **Formalia**

Grundbegriffe der Informatik

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

## Homepage der Abteilung für Methodenlehre I



INSTITUT FÜR PSYCHOLOGIE

INSTITUT | STUDIUM | FORSCHUNG | PERSONEN

[Home](#) > [Institut](#) > [Abteilungen des Ins...](#) > [Methodenlehre I : Experimentelle und Neuro...](#) > [Forschung](#) | [Lehre](#) | [Team](#)

### Methodenlehre I : Experimentelle und Neurowissenschaftliche Psychologie

#### Forschung



#### Lehre



#### Team



## C1. Programmierung und Deskriptive Statistik

- Einführung in die datenanalytische Programmierung mit R in Visual Studio Code
- Einführung in die Auswertung deskriptiver Statistiken

## C2. Analyse und Dokumentation

- Praktische Analyse empirischer Daten
- Dokumentation empirischer Studien und Analysen



## Programmierung

- Einführung in die Programmiersprache R und die Entwicklungsumgebung Visual Studio Code (VS Code).
- Grundlegende Konzepte der datenanalytischen Programmierung in R, einschließlich arithmetischer Operationen, logischer Operationen, Variablen, Datenstrukturen und Kontrollstrukturen.
- Einführung in wesentliche Werkzeuge und Pakete für die Durchführung deskriptiver statistischer Analysen mit R.

## Deskriptive Statistik

- Grundlegender Methoden der deskriptiven Statistik zur Analyse und Beschreibung quantitativer Daten.
- Praktische Übung anhand von Anwendungsbeispielen.

## Seminar und Präsenzübung

- Theoretische und praktische Einführung in die wesentlichen Konzepte der Programmierung und deskriptiven Statistik.
- Anwendung der erlernten Konzepte in praktischen Übungen, unterstützt durch Hilfestellung von Lehrpersonen und Tutor:innen während der Präsenzübungen.

## Programmierübungen

- Neben den Präsenzzeiten werden wöchentliche Vor- und Nachbereitung der Programmierübungen und Selbstkontrollfragen dringend empfohlen!
- Die Computerplätze am URZ sind außerhalb der Belegzeiten (siehe [G26.1-006](#) oder [G26.1 007](#)) öffentlich zugänglich.

## Leistungsnachweis


- Selbstständige Bearbeitung von Übungsaufgaben und deren Einreichung in Form von R-Skripten.
- Leistungsnachweis gilt als erbracht, wenn mind. 50 % der abgegebenen Leistungen als bestanden bewertet wurden.

- Termine: Dienstags
  - Gruppe 1: 11-13 Uhr in G26.1 007 (Persönlicher Laptop optional)
  - Gruppe 2: 13-15 Uhr in G22A-120 (Persönlicher Laptop notwendig)
- Kursmaterialien (Folien, Videos) auf der [Kurswebseite](#)
- Codes und tagesaktuelle Folien auf [Github](#)
- Vorherige Iteration des Kurses: [PDS \(WS 2023/24\)](#)
- Empfohlene Vorbereitung: [Vorkurs "Mathematische Grundlagen"](#)
- Ankündigungen und Fragen-Forum im [Mattermost-Channel](#)
- Abgabe und Evaluation des Leistungsnachweises über den [Moodlekurs](#)

# Termine

Datum	Einheit	Thema	Veranstaltungsform
15.10.24	R Grundlagen	(1) Einführung	Seminar
22.10.24	R Grundlagen	(2) R und Visual Studio Code	Seminar
29.10.24	R Grundlagen	(2) R und Visual Studio Code	Übung
05.11.24	R Grundlagen	(3) Vektoren, (4) Matrizen	Seminar
12.11.24	R Grundlagen	(5) Listen und Dataframes	Seminar
	<i>Leistungsnachweis 1</i>		
19.11.24	R Grundlagen	(6) Datenmanagement	Seminar
26.11.24	R Grundlagen	(2-6) R Grundlagen	Übung
03.12.24	Deskriptive Statistik	(7) Häufigkeitsverteilungen	Seminar
10.12.24	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile	Seminar
	<i>Leistungsnachweis 2</i>		
17.12.24	Deskriptive Statistik	(9) Maße der zentralen Tendenz und Variabilität	Seminar
	Weihnachtspause		
07.01.25	R; Deskriptive Statistik	(10) Kontrollstrukturen; Anwendungsbeispiel	Seminar
14.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Übung
	<i>Leistungsnachweis 3</i>		
21.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Seminar
28.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel, Q&A	Seminar


## Webseite des Kurses (Folien, Videos)



OTTO VON GUERICKE  
UNIVERSITÄT  
MAGDEBURG

INSTITUT FÜR PSYCHOLOGIE

Sitemap Impressum Kontakt

Suchbegriff 

INSTITUT | STUDIUM | FORSCHUNG | PERSONEN

DIREKTLINKS ▾

Home > Methodenlehre I > Lehre > Wintersemester ... > Programmierung und Deskriptive Stati...

### Programmierung und Deskriptive Statistik

Dieser Kurs gibt eine Einführung in die datenanalytische Programmierung und die Auswertung deskriptiver Statistiken mit [R](#) in [Visual Studio Code](#).

Nach der [Studien- und Prüfungsordnung für den BSc Psychologie \(06/2020\)](#) und dem [Modulhandbuch für den BSc Psychologie \(09/2020\)](#) entspricht dieser Kurs dem Modul [CT Computergestützte Datenanalyse](#).

Der Quarto Code der Vorlesungsfolien ist auf [github](#) verfügbar.

Zur Überprüfung der erfolgreichen Teilnahme an diesem Kurs wird ein Leistungsnachweis in zwei Teilen abgegeben.

Als weiterführende Literatur zur Programmierung mit R werden empfohlen:


- Sauer, S. (2019) Moderne Datenanalyse mit R
- Cotton, R. (2013) Learning R
- Wickham, H. (2019) Advanced R
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023) R for Data Science
- Murrell, P. (2021) R Graphics

Als Einstieg in die Deskriptive Statistik bieten sich an

- Henze, N. (2016) Deskriptive Statistik
- Fahrmeier et al. (2016) Statistik, Kapitel 1-3

### Vorlesungseinheiten

(1) Einführung



**Kontakt**

**Abteilungsleitung**

• Prof. Dr. Dirk Ostwald

[dirk.ostwald@ovgu.de](mailto:dirk.ostwald@ovgu.de)

Tel.: + 49 391 67 57370

**Abteilungsassistentz**

• Birgit Müller

[birgit.mueller@ovgu.de](mailto:birgit.mueller@ovgu.de)

Tel.: +49 391 67 58464

**Anschrift**

Otto-von-Guericke-Universität  
Magdeburg  
Institut für Psychologie  
Universitätsplatz 2  
Gebäude 24  
39106 Magdeburg

• [Anfahrt](#)

## Git-repository des Kurses (Code und Folien)

The screenshot shows the GitHub repository page for 'belindamef/progr-und-deskr-stat-25'. The repository is public and has 21 commits. The file list on the left includes folders for '1\_Einfuehrung', '2\_R\_und\_VSCode', '3\_Vektoren', '4\_Matrizen', '5\_Listen\_und\_Dataframes', '6\_Datenmanagement', '7\_Aufgabenverteilungen', '8\_Verteilungsfunktionen\_und\_Quantile', '9\_Masse\_der\_Zentralen\_Tendenz\_und\_Varia...', 'Abbildungen', 'Daten', and files for '.gitignore', 'Headetectex', 'README.md', 'R\_common.R', and 'Referenzen.bib'. The right sidebar shows the 'About' section with no description, website, or topics provided, and the 'Releases' and 'Packages' sections with no published items.

**Repository: progr-und-deskr-stat-25** (Public)

364841 · 5 days ago · 21 Commits

1 main · 1 Branch · 0 Tags

Go to file · Add file · Code

File/Folder	Description	Commit Date
1_Einfuehrung	Rephrase exercise section headings to use "Programmie...	2 weeks ago
2_R_und_VSCode	Update Programmieraufgaben und small files	last week
3_Vektoren	Add instructions and exercises for help and viewer in (3) ...	2 weeks ago
4_Matrizen	Add instructions and exercises for help and viewer in (3) ...	2 weeks ago
5_Listen_und_Dataframes	Add screenshots of list and table viewer to (5)	5 days ago
6_Datenmanagement	Small Files (6)	5 days ago
7_Aufgabenverteilungen	Add script for demonstrating VSCode Table Viewers	5 days ago
8_Verteilungsfunktionen_und_Quantile	Add (8)	5 days ago
9_Masse_der_Zentralen_Tendenz_und_Varia...	Add (9)	5 days ago
Abbildungen	Add images and data	5 days ago
Daten	Add images and data	5 days ago
.gitignore	chore: Add initial project files and configuration	last month
Headetectex	chore: Add initial project files and configuration	last month
README.md	chore: Add initial project files and configuration	last month
R_common.R	chore: Add initial project files and configuration	last month
Referenzen.bib	chore: Add initial project files and configuration	last month

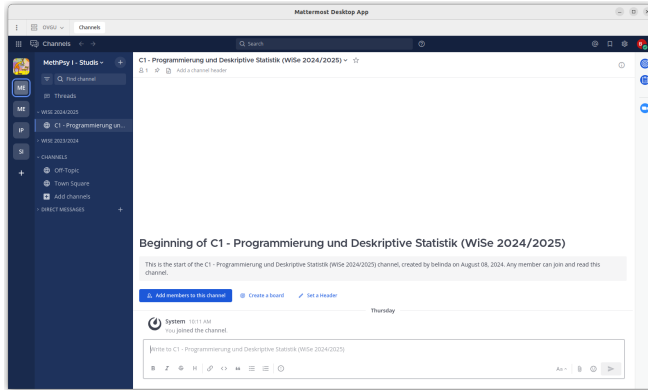
**About**  
No description, website, or topics provided.

**Releases**  
No releases published  
[Create a new release](#)

**Packages**  
No packages published  
[Publish your first package](#)

**Languages**

■ Text 99.1% ■ R 0.9%



- Registrierung und Beitritt zum **Team MethPsy I - Studis**.
- Beitritt zum **Channel C1 - Programmierung und Deskriptive Statistik (WiSe 2024/2025)**
- Zugang über **Web-App** (gleicher Link wie oben), **Desktop-App** oder **Mobile App** möglich.

## **Forum als Community Tool**

Wenn ihr Fragen habt, die für alle relevant sind oder auch andere beschäftigen könnten, stellt sie bitte im Forum. Schaut vorher, ob jemand schon eine ähnliche Frage gestellt hat. Das vermeidet Dopplungen und fördert den Austausch! Beantwortet gegenseitig eure Fragen – anderen etwas zu erklären, ist oft der beste Weg, um das eigene Verständnis zu schärfen.

## **Seid präzise und spezifisch**

Formuliert eure Frage bitte möglichst spezifisch. Statt „Ich verstehe das Thema X nicht.“ formuliert spezifische Fragen wie „Ist im dritten Punkt zum Thema X gemeint, dass...?“. Präzise Fragen führen zu klaren Antworten und helfen euch zudem, das Thema bereits beim Formulieren besser zu durchdringen. Oft lösen sich Fragen dabei auch von selbst.

## **Kontext ist Key**

Verweist bitte immer auf das Thema, die Foliennummer oder die Übungsaufgabe. Am besten direkt mit der Verlinkungsfunktion in Mattermost. So können wir die Frage schneller im richtigen Zusammenhang verstehen und beantworten.







# Mattermost Pro-Tipp: Markdown

Nutzt **Markdown in Mattermost**, um Inhalte übersichtlicher zu formatieren!

**Use Markdown**

You can also format your messages in Mattermost using Markdown to control [text styling](#), [links](#), [headings](#), [lists](#), [code blocks](#), [in-line code](#), [in-line images](#), [horizontal lines](#), [block quotes](#), [tables](#), and [math formulas](#). Markdown makes it easy to format messages: type a message as you normally would, then use formatting syntax to render the message a specific way. For a guide to using Markdown in Mattermost, [see this blog post](#).

Text Entered	How It Appears
<code>_italics_</code>	<i>italics</i>
<code>**bold**</code>	<b>bold</b>
<code>~~strikethrough~~</code>	<del>strikethrough</del>
<code>`In-line code`</code>	<code>In-line code</code>
<code>[hyperlink](http://mattermost.org)</code>	<a href="http://mattermost.org">hyperlink</a>
<code>!{embedded image}(https://travis-ci.org/mattermost/platform.svg)</code>	 build <b>passing</b>
<code>:smile: :sheep: :alien:</code>	  

**Text style**

You can use either `_` or `*` around a word or phrase to make it italic, or ``` or `**` around a word or phrase to make it bold.

**Tip**

Common formatting keyboard shortcuts are supported. Bold text by pressing `Ctrl` + `B` on Windows and Linux, or `Cmd` + `B` on Mac. Italicize text by pressing `Ctrl` + `I` on Windows or Linux, or `Cmd` + `I` on Mac.

- `*_italics_` (or `_italics_`) renders as *italics*
- `**bold**` renders as **bold**
- `***bold-italics***` renders as ***bold-italics***
- `~~strikethrough~~` renders as ~~strikethrough~~

Q & A

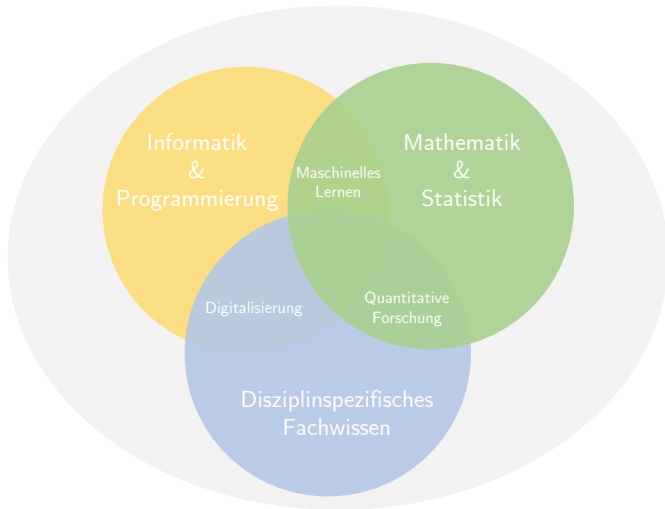
Formalia

## **Grundbegriffe der Informatik**

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

## Zentrale Komponenten der Datenwissenschaft



Formalia

## **Grundbegriffe der Informatik**

- Datenanalyse
- Informatik
- Rechnerarchitektur
- Algorithmen und Programme

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

Formalia

Grundbegriffe der Informatik

- **Datenanalyse**
- Informatik
- Rechnerarchitektur
- Algorithmen und Programme

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

- Wissenschaftliche Daten liegen heutzutage als digitale Daten vor.
- Digitale Daten werden mit Hilfe eines Computers analysiert.
- Zur Analyse von digitalen Daten schreibt man Computerprogramme.
- Diese Computerprogramme heißen Datenanalyseskripte.

1. Einlesen und Bereinigen eines digitalen Datensatzes.
2. Berechnung und Visualisierung deskriptiver Statistiken.
3. Probabilistische Modellierung und Inferenz.
4. Dokumentation und Präsentation der Ergebnisse.



## Typische Werkzeuge zur Analyse psychologischer Daten

- **R** (frei, Datenwissenschaft, Statistik, Psychologie)
- **Python** (frei, Datenwissenschaft, Anwendung)
- **Matlab** (kommerziell, Engineering, Neuroimaging)

## Altmodisch

- **SPSS** (kommerziell, Sozialwissenschaften, Psychologie)
- **JMP** (kommerziell, Biologie, Psychologie)
- **STATA** (kommerziell, Wirtschaftswissenschaften)

# Programmiersprachen Trends

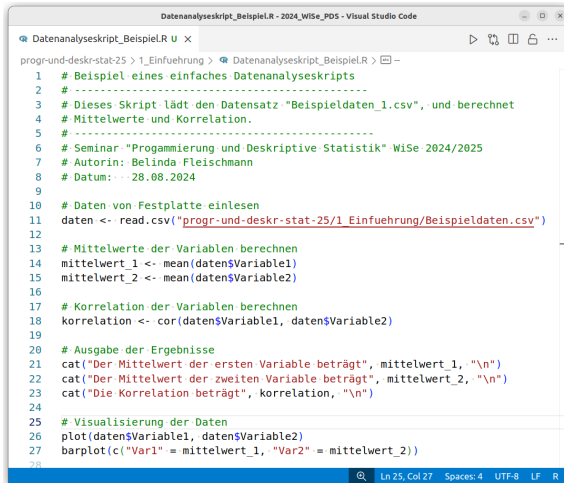
## PYPL Index (Stand: August 2024)

Worldwide, Aug 2024 :				
Rank	Change	Language	Share	1-year trend
1		Python	29.6 %	+1.7 %
2		Java	15.51 %	-0.3 %
3		JavaScript	8.38 %	-1.0 %
4		C#	6.7 %	-0.0 %
5		C/C++	6.31 %	-0.2 %
6	↑	R	4.6 %	+0.2 %
7	↓	PHP	4.35 %	-0.6 %
8		TypeScript	2.93 %	-0.1 %
9		Swift	2.76 %	+0.1 %
10	↑	Rust	2.58 %	+0.5 %
11	↓	Objective-C	2.4 %	+0.2 %
12		Go	2.14 %	+0.2 %
13		Kotlin	1.94 %	+0.2 %
14		Matlab	1.5 %	-0.0 %

- PopularitY of Programming Language
- Basierend auf GoogleSuchanfragen zu Programmiersprachentutorials

- Dokumentation aller Schritte von Rohdaten bis zur Datenvisualisierung.
- Reproduktion wissenschaftlicher Ergebnisse durch Dritte.
- Essentieller Teil wissenschaftlicher Publikationen.
- Essentieller Teil täglicher wissenschaftlicher Arbeit.

## Beispiel



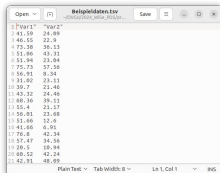
```
Datenanalyseskript_Beiispiel.R - 2024_WiSe_PDS - Visual Studio Code

Datenanalyseskript_Beiispiel.R U X
progr-und-deskr-stat-25 > 1_Einfuehrung > Datenanalyseskript_Beiispiel.R > --
1  # Beispiel eines einfachen Datenanalyseskripts
2  # -----
3  # Dieses Skript lädt den Datensatz "Beispieldaten_1.csv", und berechnet
4  # Mittelwerte und Korrelation.
5  # -----
6  # Seminar: "Programmierung und Deskriptive Statistik" WiSe 2024/2025
7  # Autorin: Belinda Fleischmann
8  # Datum: ...28.08.2024
9
10 # Daten von Festplatte einlesen
11 daten <- read.csv("progr-und-deskr-stat-25/1_Einfuehrung/Beispieldaten.csv")
12
13 # Mittelwerte der Variablen berechnen
14 mittelwert_1 <- mean(daten$Variable1)
15 mittelwert_2 <- mean(daten$Variable2)
16
17 # Korrelation der Variablen berechnen
18 korrelation <- cor(daten$Variable1, daten$Variable2)
19
20 # Ausgabe der Ergebnisse
21 cat("Der Mittelwert der ersten Variable beträgt", mittelwert_1, "\n")
22 cat("Der Mittelwert der zweiten Variable beträgt", mittelwert_2, "\n")
23 cat("Die Korrelation beträgt", korrelation, "\n")
24
25 # Visualisierung der Daten
26 plot(daten$Variable1, daten$Variable2)
27 barplot(c("Var1" = mittelwert_1, "Var2" = mittelwert_2))
28
```

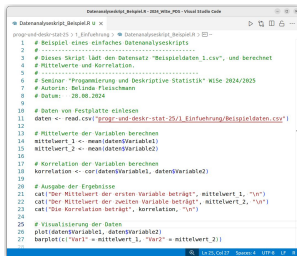
## Intuition zur Struktur der Computergestützten Datenanalyse

### Datenanalysekript

#### Datensatz



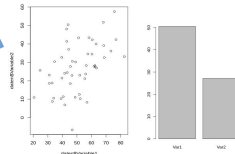
Var1	Var2
41.59	24.09
46.55	22.9
73.38	36.13
51.86	43.31
51.94	23.64
75.73	57.56
56.91	8.34
31.82	23.11
30.7	21.46
43.32	24.46
66.36	39.11
55.4	21.17
56.81	23.68
51.66	12.6
41.66	6.95
76.8	42.34
57.47	34.56
28.5	16.94
60.52	42.24
42.91	48.89



```
Datenanalysekript_Beiispiel.R u x
# Datenanalysekript_Beiispiel.R - 2024_WiSe_K25 - Visual Studio Code
# prog-and-deskr-stat-25/1_Einfuehrung > Datenanalysekript_Beiispiel.R >
1 # Beispiel eines einfachen Datenanalysekripts
2 # .....
3 # Dieses Skript lädt den Datensatz "Beispieldaten_1.csv", und berechnet
4 # Mittelwerte und Korrelation.
5 # .....
6 # Seminar "Programmierung und Deskriptive Statistik" WiSe 2024/2025
7 # Autorin: Belinda Fleischmann
8 # Datum: 28.08.2024
9
10 # Daten von Festplatte einlesen
11 daten <- read.csv("prog-and-deskr-stat-25/1_Einfuehrung/Beispieldaten.csv")
12
13 # Mittelwerte der Variablen berechnen
14 mittelwert_1 <- mean(daten$Variable1)
15 mittelwert_2 <- mean(daten$Variable2)
16
17 # Korrelation der Variablen berechnen
18 korrelation <- cor(daten$Variable1, daten$Variable2)
19
20 # Ausgabe der Ergebnisse
21 cat("Der Mittelwert der ersten Variable beträgt", mittelwert_1, "\n")
22 cat("Der Mittelwert der zweiten Variable beträgt", mittelwert_2, "\n")
23 cat("Die Korrelation beträgt", korrelation, "\n")
24
25 # Visualisierung der Daten
26 plot(daten$Variable1, daten$Variable2)
27 barplot(c("Var1" = mittelwert_1, "Var2" = mittelwert_2))
28
```

Der Mittelwert der ersten Variable beträgt 50.516  
Der Mittelwert der zweiten Variable beträgt 27.1598  
Die Korrelation beträgt 0.4851987

#### Ergebnisse als Text



#### Ergebnisse als Visualisierungen

- Die Digitalisierung betrifft insbesondere auch die Wissenschaft.
- Forschungsdatenmanagement ist eine akute Herausforderung.
- Programmierung als zentrales Handwerkszeug wissenschaftlicher Arbeit.
- Informatikkenntnisse sind in der Arbeitswelt unverzichtbar.
- Dies gilt auch für Psychotherapeut:innen (z.B. Online-Intervention).

Formalia

Grundbegriffe der Informatik

- Datenanalyse
- **Informatik**
- Rechnerarchitektur
- Algorithmen und Programme

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

## Informatik (engl. Computer Science)

Bei der Informatik handelt es sich um die Wissenschaft von der systematischen Darstellung, Speicherung, Verarbeitung und Übertragung von Informationen, wobei besonders die automatische Verarbeitung mit Computern betrachtet wird. Sie ist zugleich Grundlagen- und Formalwissenschaft als auch Ingenieurdisziplin.

*Wikipedia*

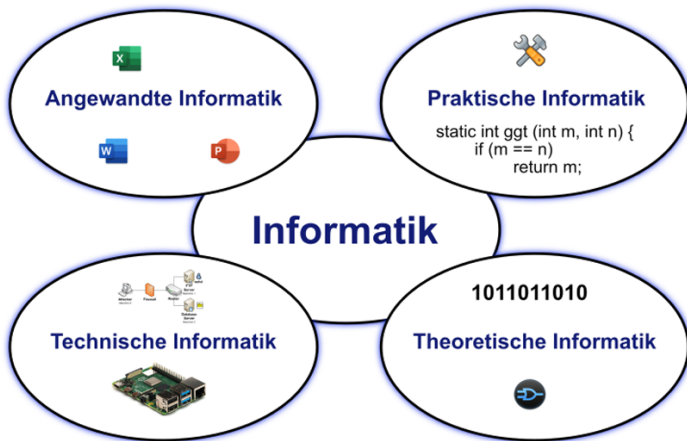


## Computer

- Maschinen zum Datenspeichern und Ausführen einfacher Datenoperationen.
- Einfache Operationen mit extrem hoher Geschwindigkeit.
- Universalität durch Speicherung von Daten und Programmen.

## Algorithmen und Programme

- *Programme* sind in einer *Programmiersprache* verfasste *Algorithmen*.
- Algorithmen sind Folgen von Anweisungen durchzuführender Operationen.
- Bei Algorithmen unterscheidet man
  - Beschreibung (Kochrezept, IKEA Bauanleitung, R Skript)
  - Anweisungen ("Mehl und Wasser vermengen",  $o \leftarrow -, x = c(1,2,3)$ )
  - Durchführung (Kochvorgang, Zusammenbau, R Skript laufen lassen)



*Hattenhauer (2020) Informatik*

## Technische Informatik

- Mikroprozessortechnik, Rechnerarchitektur, Netzwerktechnik

## Theoretische Informatik

- Automatentheorie, Berechenbarkeitstheorie, Komplexitätstheorie

## Praktische Informatik

- Programmierung, Algorithmen, Datenbanken

## Angewandte Informatik

- Anwendungssoftware, Human-Computer-Interaction, Informatik und Gesellschaft

## Maschinelles Lernen und Künstliche Intelligenz

- Datenanalyse aus Sicht der Informatik

## Computervisualistik

- Bilderkennung und Bildsynthese, Virtuelle Realität, Augmented Reality

## Computerlinguistik

- Spracherkennung und Sprachsynthese

## Bioinformatik

- Lebenswissenschaften, Genomik, Bildgebende Verfahren der Medizin

Formalia

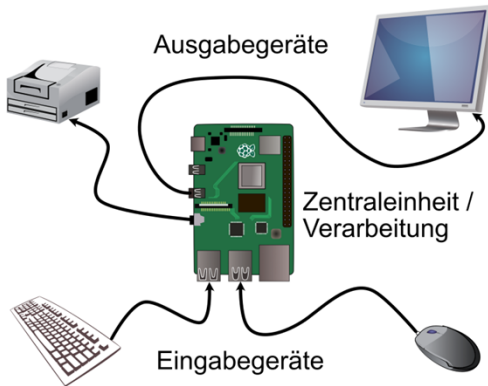
Grundbegriffe der Informatik

- Datenanalyse
- Informatik
- **Rechnerarchitektur**
- Algorithmen und Programme

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

**EVA-Prinzip:** Eingabe → Verarbeitung → Ausgabe



*Hattenhauer (2020) Informatik*

# Zentraleinheit eines Computers

Auch Hauptplatine, Motherboard oder Mainboard genannt

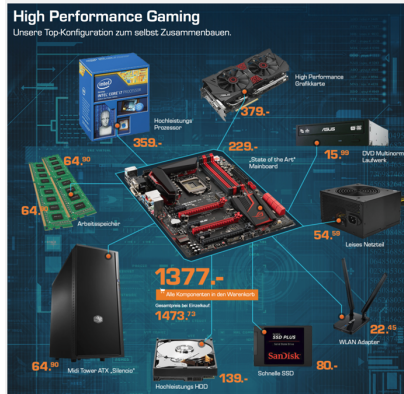


Figure 1: Hattenhauer (2020) Informatik

Hattenhauer (2020) Informatik

# Zentraleinheit eines Computers

---

## CPU (Central Processing Unit/Mikroprozessor)

- Rechenwerk, Steuerwerk, und Leitwerk des Systems
- Cache (flüchtiger schneller Speicher)
- Intel(R) Core(TM) i5-7300HQ CPU @ 2.50GHz

## RAM (Random Access Memory)

- Temporärer, flüchtiger Arbeitsspeicher des Systems
- Begrenzt, z.B. 16 GB

## Massenspeicher

- Stationärer Speicher des Systems
- SSD (Solid State Drive), Cloudspeicher

## GPU (Graphical Processing Unit)

- Leistungsstarke, speziell für Visualisierung optimierte Prozessoren
- Unterstützung der CPU in manchen Anwendungen, z.b. Neuronale Netze



Abstraktion eines Rechensystems mit wohldefinierten Komponenten und Datenflüsse.

Rechner := Steuerwerk, Rechenwerk, Speicher, Eingabewerk, Ausgabewerk.

Zentrale Eigenschaften

- Struktur des Rechners unabhängig von dem zu bearbeitenden Problem.
- Daten, Programme, Zwischen- und Endergebnisse liegen im gleichen Speicher.
- Speicher ist in gleichgroße nummerierte (adressierte) Zellen unterteilt.
- Über die Adresse einer Speicherzelle kann deren Inhalt abgerufen verändert werden.
- Ein Programm ist eine Reihe von Befehlen.
- Aufeinanderfolgende Befehle eines Programms liegen in benachbarten Speicherzellen und werden entsprechend nacheinander aufgerufen.

→ **Die Architektur eines Rechners impliziert das Grundprinzip der Programmierung:  
Befehle werden streng sequentiell abgearbeitet.**

Formalia

Grundbegriffe der Informatik

- Datenanalyse
- Informatik
- Rechnerarchitektur
- **Algorithmen und Programme**

R und Visual Studio Code

Aufgaben und Selbstkontrollfragen

# Vom Realweltproblem zum Programm

---

## Realweltproblem

- Das Problem, das mithilfe eines Computers gelöst werden soll.
- z.B. Auswertung von Fragebogendaten einer psychologischen Studie.

## Problemspezifikation

- Genaue sprachliche Fassung des Realweltproblems.
- z.B. Methodenteil einer wissenschaftlichen Publikation.

## Algorithmus

- Folge von Anweisungen zur Lösung des Problems.
- z.B. Dateneinlesen, deskriptive Statistiken berechnen, T-Test durchführen.

## Programm

- Ein Algorithmus, der von einem Computer ausgeführt werden kann.
- Eine in einer Programmiersprache verfasste Textdatei.

## Definition (Algorithmus)

Ein *Algorithmus* ist eine Folge von Anweisungen, um aus gewissen Eingabedaten bestimmte Ausgabedaten herzuleiten, wobei folgende Bedingungen erfüllt sein müssen

- *Fintheit*. Die Anweisungsfolge muss in einem endlichen Text vollständig beschrieben sein.
- *Effektivität*. Jede Anweisung muss tatsächlich ausführbar sein.
- *Terminierung*. Der Algorithmus endet nach endlich vielen Anweisungen.
- *Determiniertheit*. Der Ablauf des Algorithmus ist zu jedem Punkt fest vorgeschrieben.

Wenn  $E$  die Menge der zulässigen Eingabedaten und  $A$  die Menge der zulässigen Ausgabedaten bezeichnet, dann ist ein Algorithmus eine Funktion

$$f : E \rightarrow A, e \mapsto f(e) \quad (1)$$

Umgekehrt heißen Funktionen, die durch einen Algorithmus beschrieben werden können, *berechenbare Funktionen*.

### Bemerkung

- Effektivität sollte nicht mit Effizienz verwechselt werden.

## Eine Programmiersprache

- ... bestimmt die Regeln, denen ein Programm gehorchen muss.
- ... definiert eine Syntax, also Vokabular und Programmaufbau.
- ... definiert Semantik, also die Bedeutung der erlaubten Anweisungen.

```
#if [ -z "$USER_NAME" -o -z "$USER_TYPE" -o -z "$GROUP" ]
if [ -z "$USER_NAME" -o -z "$USER_TYPE" ]
then
#     echo "Please set the user name, type and group"
    echo "Please set the user name and type"
    exit 1
fi

# generate a random password
# -y: include special characters
# -n: include numbers
# -l: one generated passwords per line
#pwgen -y 15 -n 5 -l
echo "Propositions for random passwords to use in next step:"
pwgen -s -n -l 15 5

# add the user
# requires password to be given via input
adduser --firstuid 1000 --lastuid 9999 --no-create-home ${USER_NAME}
```

## Maschinensprache

- Elementare Operationsbefehle (z.B. Speichern, Vergleichen, Addieren)
- Elementare Operationsbefehle werden als Binärzahlen kodiert

Addiere Inhalt R1 zu Inhalt R2     $\Rightarrow$  1001 0010

Erhöhe Inhalt R1 um 1                 $\Rightarrow$  1001 0110

Übertrage Inhalt R1 nach R3         $\Rightarrow$  0010 0011

- Programme in Maschinensprache heißen *Maschinenprogramme*.
- De facto führt ein Computer nur Maschinenprogramme aus.
- Für Menschen ist die Programmierung in Maschinensprache mühselig.

## Höhere Programmiersprache

- An die menschliche Sprache angelehnte Wörter und Sätze
- Interpreter oder Compiler übersetzen Programme in Maschinensprache
- R, Python, Matlab, C++, Java, FORTRAN, COBOL,...

# Generationen von Programmiersprachen

---

## 1. Generation (1GL)

- Maschinensprachen
- 10110000 01100001 (hexadezimaler Darstellung des Ausdrucks "B0 61")

## 2. Generation (2GL)

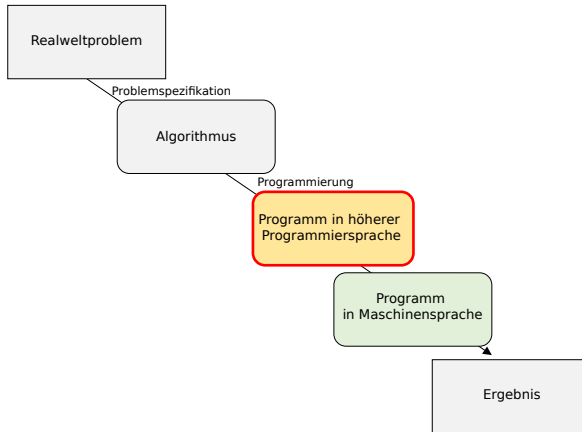
- Assemblersprachen ab 1950, erste Form der symbolischen Programmierung
- Bspw. "MOV AI, 61H" # Intel-Prozessor-spezifische Sprache

## 3. Generation (3GL)

- Höhere Programmiersprachen ab 1970 wie FORTRAN, C, C++, Java
- Programmierfreundlich, prozessor-unabhängig

## 4. Generation (4GL)

- Höhere Programmiersprachen ab 1980 wie Python, Matlab, R
- Codeoverhead Minimisierung, Automation, Flexibilität, Multiparadigmatisch





## Imperative Programmierung

- Problemlösungsweg wird als Folge von *Anweisungen (Befehlen)* vorgegeben.
- Befehle verarbeiten Daten, die mithilfe von *Variablen* adressiert werden.
  - **Prozedurale imperative Programmierung**
    - Daten und sie manipulierende Befehle werden separat behandelt.
    - Prozeduren (Funktionen) bilden das zentrale Strukturkonzept.
  - **Objektorientierte imperative Programmierung**
    - Daten und manipulierende Befehle werden als *Objekte* zusammengefasst.
    - Objekte bilden das zentrale Strukturkonzept.
- Praktisch liegen oft Mischformen vor.

## Kompilierte Programmiersprachen

- Gesamter Quellcode wird *vor der Ausführung* in Maschinensprache übersetzt.
- Das Übersetzungsprogramm heißt *Compiler*.
- Der übersetzte Maschinencode wird vom Prozessor ausgeführt.
- Das ausführbare Programm wird nicht übersetzt und läuft schnell.
- Bei Änderungen des Quellcodes muss neu kompiliert werden.
- Beispiele für kompilierte Sprachen sind Java, C, C++.

## Interpretierte Programmiersprachen

- Quellcode wird *während der Ausführung* in maschinennahe Sprache übersetzt.
- Das Ausführungsprogramm heißt *Interpreter*.
- Das Programm läuft aufgrund der Interpretation langsamer.
- Bei Änderungen des Quellcodes muss nicht neu interpretiert werden.
- Beispiele für interpretierte Sprachen sind Python und R.

Formalia

Grundbegriffe der Informatik

**R und Visual Studio Code**

Aufgaben und Selbstkontrollfragen

# Was ist R?

---

- Eine Programmiersprache und ein Softwarepaket.
- Entwickelt von Ihaka and Gentleman (1996).
- Freier Dialekt der proprietären Software S (Becker, Chambers, and Wilks (1988)).
- Weiterentwickelt und gepflegt durch [R Core Team](#) und [R Foundation](#)
- Interpretierte imperative 4GL Sprache.
- Optimiert und populär für statistische Datenanalysen.
- Große Community mit etwa 20.000 beigetragenen [R Paketen](#) (Erweiterungen)
- Evolviert und konservativ im Kern, konsistent und progressiv in R Paketen.

# Wie bekomme ich R?

Über [cran.r-project.org](https://cran.r-project.org) die geeignete Version herunterladen und installieren.



CRAW  
Mimms  
What's new?  
Search  
[CRAN Team](#)  
  
About R  
[R Homepage](#)  
[The R Journal](#)  
  
Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Task Views](#)  
[Other](#)  
  
Documentation  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

<b>Download and Install R</b>
Precompiled binary distributions of the base system and contributed packages, <b>Windows</b> and <b>Mac</b> users most likely want one of these versions of R: <ul style="list-style-type: none"><li>• <a href="#">Download R for Linux (Debian, Fedora/Redhat, Ubuntu)</a></li><li>• <a href="#">Download R for macOS</a></li><li>• <a href="#">Download R for Windows</a></li></ul>
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.
<b>Source Code for all Platforms</b>
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it! <ul style="list-style-type: none"><li>• The latest release (2022-06-23, Funny-Looking Kid) <a href="#">R-4.2.1.tar.gz</a>, read <a href="#">what's new</a> in the latest version.</li><li>• Sources of <a href="#">R alpha and beta releases</a> (daily snapshots, created only in time periods before a planned release).</li><li>• Daily snapshots of current patched and development versions are <a href="#">available here</a>. Please read about <a href="#">new features</a> and <a href="#">bug fixes</a> before filing corresponding feature requests or bug reports.</li><li>• Source code of older versions of R is <a href="#">available here</a>.</li><li>• Contributed extension <a href="#">packages</a></li></ul>
<b>Questions About R</b>
<ul style="list-style-type: none"><li>• If you have questions about R like how to download and install the software, or what the license terms are, please read our <a href="#">answers to frequently asked questions</a> before you send an email.</li></ul>

## What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to minimize network load.

## Submitting to CRAN

To "submit" a package to CRAN, check that your [submission](#) meets the [CRAN Repository Policy](#) and then use the [web form](#).

If this fails, send an email to [CRAN-submissions@R-project.org](mailto:CRAN-submissions@R-project.org) following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

# Was können wir mit R machen?

---

## Was kann R?

- Datensätze laden, manipulieren, und speichern.
- Eine Vielzahl von Berechnungen an verschiedenen Datenstrukturen durchführen.
- Eine Vielzahl statistischer Analysemethoden auf Daten anwenden.
- Datenanalyseskripte schreiben und Abbildungen generieren.
- Präsentationen [RMarkdown](#) und Bücher [RBookdown](#) erstellen.
- Wissenschaftliche Berichte mit [Quarto](#) erstellen.

## Was kann R (bisher) nicht so gut?

- In einer ansprechenden Umgebung programmieren ( $\Rightarrow$  Visual Studio Code).
- Scientific Computing ( $\Rightarrow$  Python, Matlab, Julia).
- Psychologische Experimente programmieren ( $\Rightarrow$  Python, Matlab)

# Wie bekomme ich Hilfe zu R?

- Während der Programmierung und bei bekanntem Funktionsnamen über die Kommandozeile:

```
?mean          # Zeigt Hilfe zu der Funktion "mean()"
help(mean)      # Zeigt Hilfe zu der Funktion "mean()"
browseVignettes() # Zeigt Vignetten aller installierten Pakete im Browser
browseVignettes("knitr") # Zeigt Vignetten des Paktes "knitr"
```

- Googlen
- [stackoverflow.com](https://stackoverflow.com)
- [r-project.org/help.html](https://r-project.org/help.html)
- [rseek.org](https://rseek.org)
- [rstudio.com/resources/cheatsheets](https://rstudio.com/resources/cheatsheets)
- [r-bloggers.com](https://r-bloggers.com)

# Was ist Visual Studio Code (VSCode)?

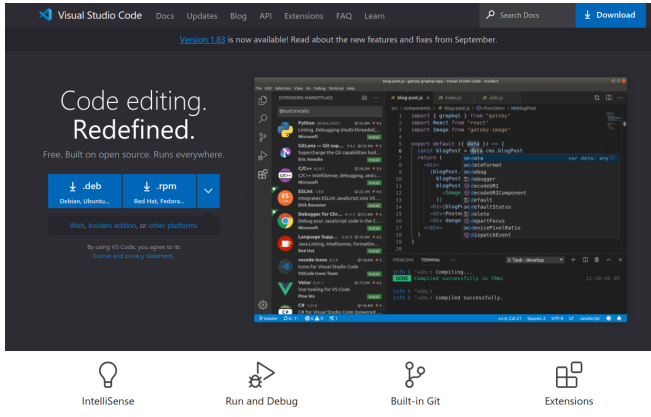
---

- VSCode ist ein kostenloser Quelltext-Editor von Microsoft.
- VSCode ist eine Softwareentwicklungsumgebung (Integrated Development Environment, IDE)
- Seit 2015 für Windows, macOS und Linux verfügbar.
- Seit 2018 ist VSCode der beliebteste Editor laut jährlicher stackoverflow Umfragen  
([Aktuellste aus 2024](#)).
- Ein Microsoftprodukt ist damit auch der beliebteste Editor der Linuxwelt.
- Über Extensions kann VSCode als IDE für beliebige Sprachen genutzt werden.
- Zum Beispiel funktioniert VSCode als IDE für R, Python, Julia, Shell, Quarto, etc.
- VSCode ist Community-based und sehr konfigurierbar.
- VSCode ist über Microsoft's GitHub über Endgeräte synchronisierbar.



## Wie bekomme ich VSCode?

Über <https://code.visualstudio.com> herunterladen und installieren.



# Wie benutze ich VSCode?

Online Dokumentation: <https://code.visualstudio.com/docs>

The screenshot shows the Visual Studio Code documentation website. The header includes the Visual Studio Code logo, navigation links (Docs, Updates, Blog, API, Extensions, FAQ, Learn), a search bar, and a 'Download' button. A banner below the header announces 'Version 1.82 is now available! Read about the new features and fixes from September.' The main content area is titled 'Getting Started' and describes Visual Studio Code as a lightweight but powerful source code editor. It includes a list of languages supported (Node.js, JavaScript, TypeScript, Python, Java, C++, C#, Docker, Data Science, Azure, Remote, Dev Containers) and a list of actions (Subscribe, Ask questions, Follow @code, Request features, Report issues, Watch videos). The 'Intelligent Code Completion' section highlights the IntelliSense feature, showing a code snippet with a dropdown menu of suggestions for the 'server.get' method.

Visual Studio Code

Docs Updates Blog API Extensions FAQ Learn

Search Docs Download

Version 1.82 is now available! Read about the new features and fixes from September.

## Getting Started

Visual Studio Code is a lightweight but powerful source code editor which runs on your desktop and is available for Windows, macOS and Linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages and runtimes (such as C++, C#, Java, Python, PHP, Go, .NET). Begin your journey with VS Code with these [introductory videos](#).

### Visual Studio Code in Action

```
4 var server = express();
5 server.use(bodyParser.json);
6
7 server.get
8   get (property) Application.get: ((name: string)...
9   getMaxListeners
10  arguments
11  engine
12  length
13  merge
14  purge
15  settings
16  toString
17  defaultConfiguration
18
```

#### Intelligent Code Completion

Code smarter with **IntelliSense** - completions for variables, methods, and imported modules.

Formalia

Grundbegriffe der Informatik

R und Visual Studio Code

**Aufgaben und Selbstkontrollfragen**

# Aufgaben bis nächste Woche

---

1. R installieren.
2. VSCode installieren.
3. VSCode für R startklar machen (Anleitung [hier](#)).
4. Selbstkontrollfragen bearbeiten.
5. Mit dem eigenen oder Labor-Rechner vertraut machen.
  - Wie viel Festplatten- und Arbeitsspeicher besitzt der Rechner?
  - Welches Betriebssystem in welcher Version wird verwendet?
  - Erstellt einen Ordner für alle Materialien und Übungen dieses Kurses (z.B. "PDS\_2024") an einem Speicherort eurer Wahl. Vermeidet Umlaute und Leerzeichen.
  - Wie lautet die vollständige Speicheradresse (Pfad) zu diesem Ordner?
  - Ladet die Dateien "Beispiel\_Datenanalyseskript.R" und "Beispieldaten.csv" herunter und speichert sie in diesem Ordner.
  - Mit welchen Programmen könnt ihr diese Dateien jeweils öffnen und mit welchen nicht?

1. Gebe die typische Struktur einer computergestützten Datenanalyse wieder.
2. Erläutere den Begriff "Datenanalyseskript".
3. Definiere den Begriff "Informatik".
4. Erläutere Sie die Akronyme CPU, RAM, SSD, und GPU.
5. Nenne die wesentlichen Aspekte der Von-Neumann Rechnerarchitektur.
6. Definiere den Begriff Algorithmus.
7. Erläutere den Zusammenhang von Algorithmen und Programmen.
8. Was bezeichnen Syntax und Semantik einer Programmiersprache?
9. Differenziere die Begriffe "Maschinensprache" und "höhere Programmiersprache".
10. Skizziere Prinzipien der prozeduralen und objektorientierten imperativen Programmierung.
11. Skizziere die Entwicklung der Programmiersprachen der ersten bis vierten Generation.
12. Differenziere die Begriffe der kompilierten und der interpretierten Programmiersprachen.

- Becker, Richard A., John M. Chambers, and Allen Reeve Wilks. 1988. *The New S Language: A Programming Environment for Data Analysis and Graphics*. Reprint. London: Chapman & Hall.
- Ihaka, Ross, and Robert Gentleman. 1996. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5 (3): 2999–2314.