

Data Science & Machine Learning

Dieter Greipl

2022-01-21

Contents

1	Daten	3
1.1	Der Iris-Datensatz	3
1.2	Datentypen	6
1.3	Skalenniveaus	6
2	Daten	8
2.1	Der Iris-Datensatz	8
2.2	Datentypen	11
2.3	Skalenniveaus	11
3	Iris-Datensatz	13
3.1	Einführung	13
3.2	Datensatz	13
4	Skalenniveaus	16
5	Blocks	18
5.1	Equations	18
5.2	Theorems and proofs	18
5.3	Callout blocks	18
Teil B		20
6	Sharing your book	20
6.1	Publishing	20
6.2	404 pages	20
6.3	Metadata for sharing	20
7	Footnotes and citations	21
7.1	Footnotes	21
7.2	Citations	21
8	Template	22

9	Template	23
----------	-----------------	-----------

Chapter 1

Daten

1.1 Der Iris-Datensatz

Der Iris-Datensatz enthält Messungen von jeweils 50 Blüten zu drei verschiedenen Lilien-Arten (*setosa*, *versicolor*, *virginica*)

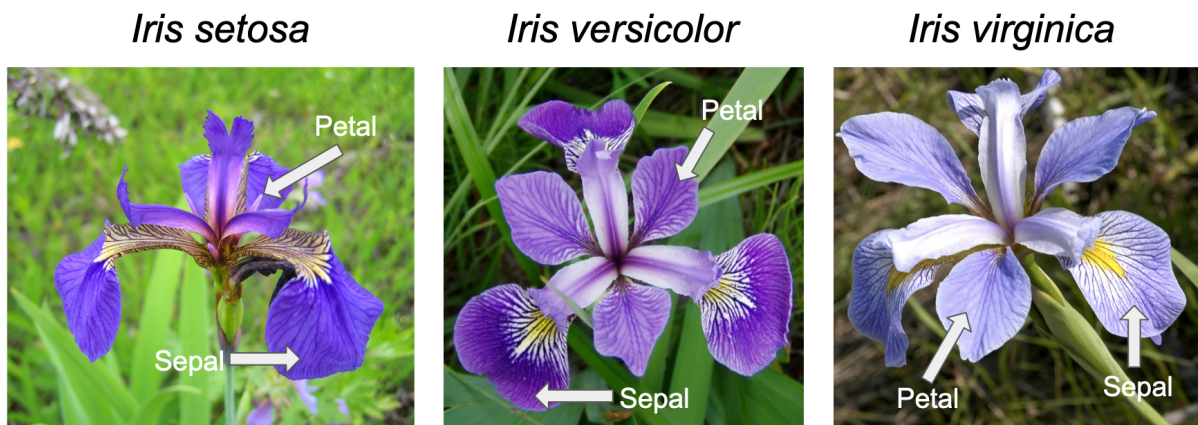


Figure 1.1: Download

Gemessen werden pro Blüte in cm

- die Länge und Breite des Kronblattes (Petalum, petal) sowie
- die Länge und Breite des Kelchblattes (Sepalum, sepal)

1.1.1 Datensatz

Folgender - in der Community wohlbekannter - Datensatz liegt uns vor (Sie finden die Daten hier).

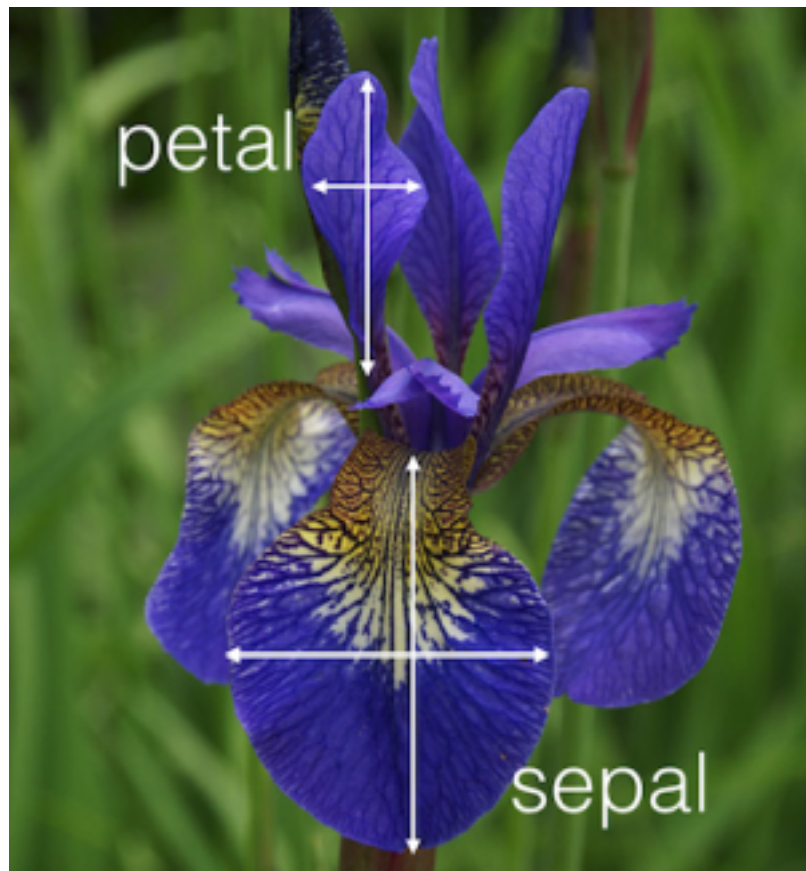


Figure 1.2: image (190)

	A	B	C	D	E	F
1	sepal_len	sepal_wid	petal_len	petal_wid	class	
2	7,2	3,6	6,1	2,5	virginica	
3	5	3,6	1,4	0,2	setosa	
4	4,4	3	1,3	0,2	setosa	
5	4,7	3,2	1,6	0,2	setosa	
6	6,7	3,1	4,7	1,5	versicolor	
7	5,7	3,8	1,7	0,3	setosa	
8	6,9	3,1	5,4	2,1	virginica	
9	6,1	2,8	4,7	1,2	versicolor	
10	6,3	3,3	6	2,5	virginica	
11	6,5	3	5,2	2	virginica	
12	5,9	3	5,1	1,8	virginica	
13	6,3	2,5	5	1,9	virginica	
14	6,9	3,1	4,9	1,5	versicolor	
15	5,4	2,7	1,5	0,2	setosa	

Figure 1.3: Iris-Datensatz

1.2 Datentypen

1.2.1 Elementare Datentypen

1.2.1.1 Zahlen

1.2.1.2 Strings

1.2.1.3 Logische Werte

1.2.1.4 Elementare Datentypen in Python

1.2.2 Komplexe Datentypen

1.2.2.1 Datum

1.2.2.2 Uhrzeit

1.2.2.3 Bilder

1.2.2.4 Komplexe Datentypen in Python

1.3 Skalenniveaus


1.3.1 Überblick

Scale	Operations	Description	Statistics	Example
Nominal	$=, \neq$	values have no natural order; they describe unordered categories	Mode (Modus)	München, Hamburg, Essen
Ordinal	$<, >$	values have a defined order; difference of values is undefined or has no clear or meaningful definition	Median	Schulnoten, Tabellenplatz in der Bundesliga
Interval	$+, -$	differences of values have the same meaning; adding provides useful results; zero point is not naturally/globally defined	Mean	Temperatur
Ratio	$\cdot, /$	zero point is naturally defined	(Generalized) Mean	Alter

Bemerkungen:

1. Skalenniveaus sind nicht immer klar zuzuordnen.
2. Aus nominalen Datenskalen lassen sich stets *künstliche Ordnungen* (und damit ordinale Datenskalen) definieren.
3. Bilden sie keine Mittelwerte auf Daten mit ordinalen Datenskalen!
4. Nominale und ordinale Datenskalen heißen auch *kategorisch* oder *qualitativ*.
5. Intervall und Ratio-Datenskalen heißen auch *metrisch*.

Ergänzend: Die fünf Skalenniveaus: Einfach und verständlich erklärt (statistikpsychologie.de)



	A	B	C	D	E
1	sepal_len	sepal_wid	petal_len	petal_wid	class
2	7,2	3,6	6,1	2,5	virginica
3	5	3,6	1,4	0,2	setosa
4	4,4	3	1,3	0,2	setosa

Figure 1.4: Skalenniveaus bei Iris

1.3.2 Skalenniveaus im Iris-Datensatz

1.3.2.0.1 Von Nominal zu Ordinal Wir werden später folgende eindeutige Zuordnung treffen:

Nominaler Wert	Ordinaler Wert
setosa	0
versicolor	1
virginica	2

Chapter 2

Daten

2.1 Der Iris-Datensatz

Der Iris-Datensatz enthält Messungen von jeweils 50 Blüten zu drei verschiedenen Lilien-Arten (setosa, versicolor, virginica)

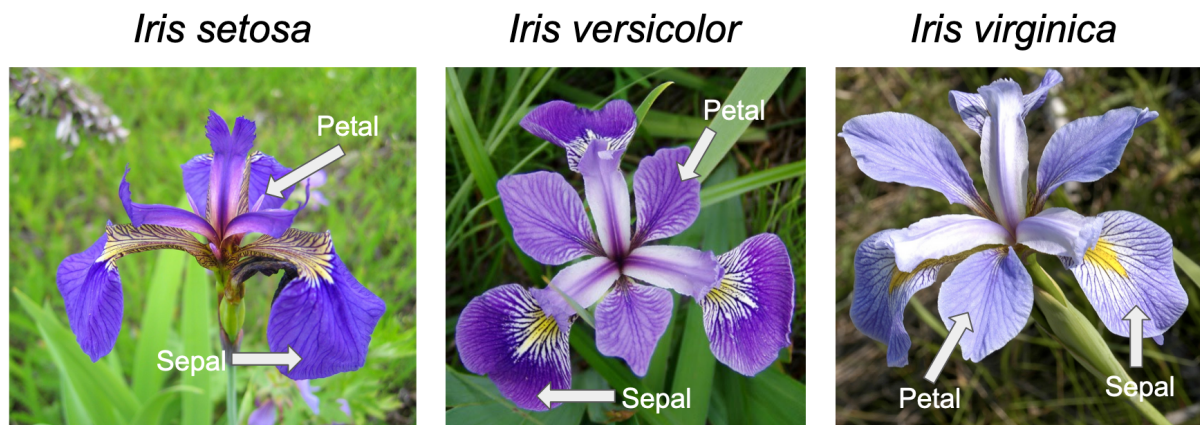


Figure 2.1: Download

Gemessen werden pro Blüte in cm

- die Länge und Breite des Kronblattes (Petalum, petal) sowie
- die Länge und Breite des Kelchblattes (Sepalum, sepal)

2.1.1 Datensatz

Folgender - in der Community wohlbekannter - Datensatz liegt uns vor (Sie finden die Daten hier).

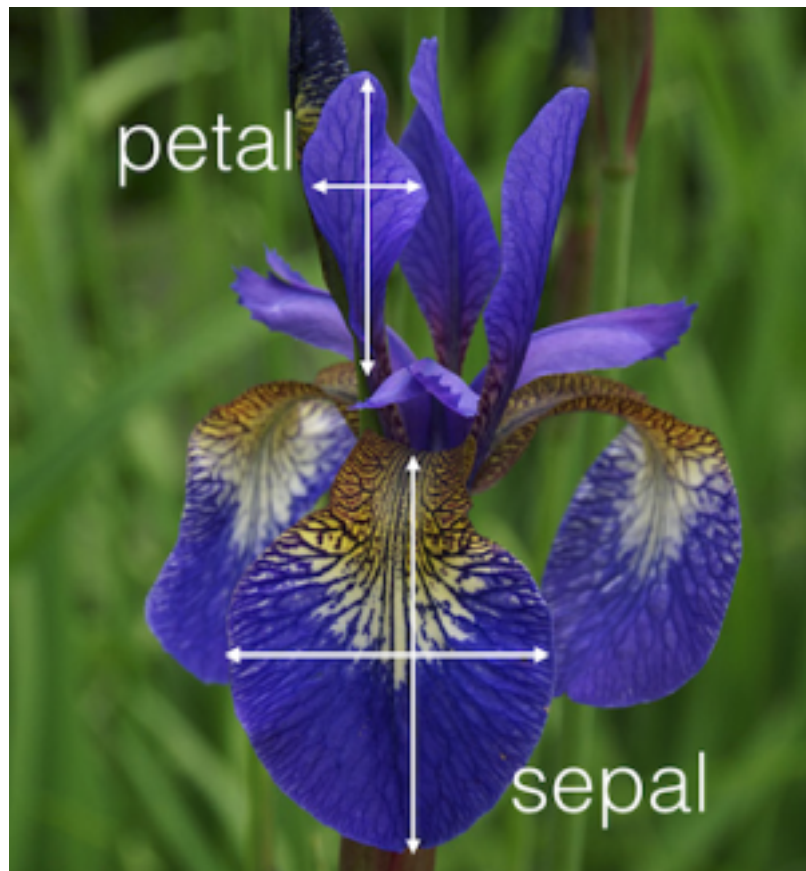


Figure 2.2: image (190)

	A	B	C	D	E	F
1	sepal_len	sepal_wid	petal_len	petal_wid	class	
2	7,2	3,6	6,1	2,5	virginica	
3	5	3,6	1,4	0,2	setosa	
4	4,4	3	1,3	0,2	setosa	
5	4,7	3,2	1,6	0,2	setosa	
6	6,7	3,1	4,7	1,5	versicolor	
7	5,7	3,8	1,7	0,3	setosa	
8	6,9	3,1	5,4	2,1	virginica	
9	6,1	2,8	4,7	1,2	versicolor	
10	6,3	3,3	6	2,5	virginica	
11	6,5	3	5,2	2	virginica	
12	5,9	3	5,1	1,8	virginica	
13	6,3	2,5	5	1,9	virginica	
14	6,9	3,1	4,9	1,5	versicolor	
15	5,4	2,7	1,5	0,2	setosa	

Figure 2.3: Iris-Datensatz

2.2 Datentypen

2.2.1 Elementare Datentypen

2.2.1.1 Zahlen

2.2.1.2 Strings

2.2.1.3 Logische Werte

2.2.1.4 Elementare Datentypen in Python

2.2.2 Komplexe Datentypen

2.2.2.1 Datum

2.2.2.2 Uhrzeit

2.2.2.3 Bilder

2.2.2.4 Komplexe Datentypen in Python

2.3 Skalenniveaus


2.3.1 Überblick

Scale	Operations	Description	Statistics	Example
Nominal	$=, \neq$	values have no natural order; they describe unordered categories	Mode (Modus)	München, Hamburg, Essen
Ordinal	$<, >$	values have a defined order; difference of values is undefined or has no clear or meaningful definition	Median	Schulnoten, Tabellenplatz in der Bundesliga
Interval	$+, -$	differences of values have the same meaning; adding provides useful results; zero point is not naturally/globally defined	Mean	Temperatur
Ratio	$\cdot, /$	zero point is naturally defined	(Generalized) Mean	Alter

Bemerkungen:

1. Skalenniveaus sind nicht immer klar zuzuordnen.
2. Aus nominalen Datenskalen lassen sich stets *künstliche Ordnungen* (und damit ordinale Datenskalen) definieren.
3. Bilden sie keine Mittelwerte auf Daten mit ordinalen Datenskalen!
4. Nominale und ordinale Datenskalen heißen auch *kategorisch* oder *qualitativ*.
5. Intervall und Ratio-Datenskalen heißen auch *metrisch*.

Ergänzend: Die fünf Skalenniveaus: Einfach und verständlich erklärt (statistikpsychologie.de)



	A	B	C	D	E
1	sepal_len	sepal_wid	petal_len	petal_wid	class
2	7,2	3,6	6,1	2,5	virginica
3	5	3,6	1,4	0,2	setosa
4	4,4	3	1,3	0,2	setosa

Figure 2.4: Skalenniveaus bei Iris

2.3.2 Skalenniveaus im Iris-Datensatz

2.3.2.0.1 Von Nominal zu Ordinal Wir werden später folgende eindeutige Zuordnung treffen:

Nominaler Wert	Ordinaler Wert
setosa	0
versicolor	1
virginica	2

Chapter 3

Iris-Datensatz

3.1 Einführung

Der Iris-Datensatz enthält Messungen von jeweils 50 Blüten zu drei verschiedenen Lilien-Arten (*setosa*, *versicolor*, *virginica*)

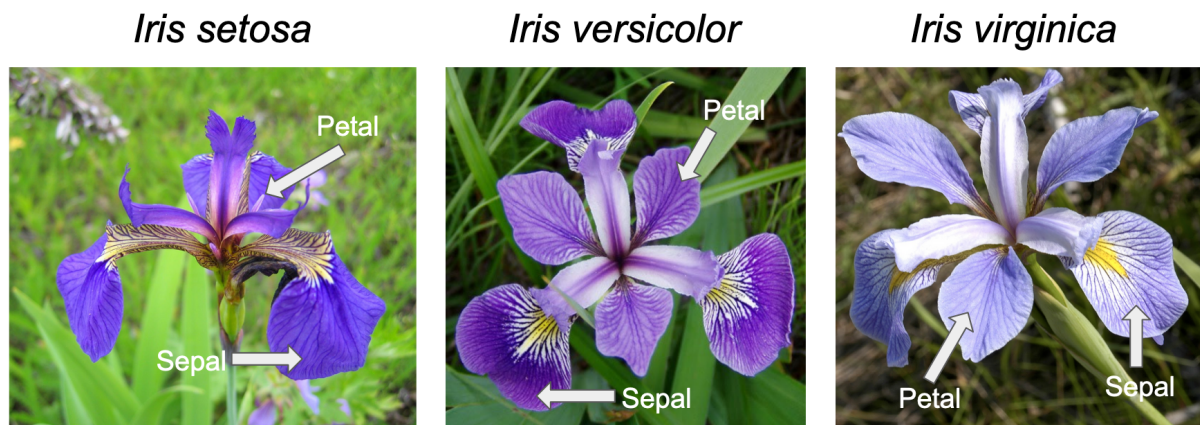


Figure 3.1: Iris-Arten¹

Gemessen werden pro Blüte in cm

- die Länge und Breite des Kronblattes (Petalum, petal) sowie
- die Länge und Breite des Kelchblattes (Sepalum, sepal)

3.2 Datensatz

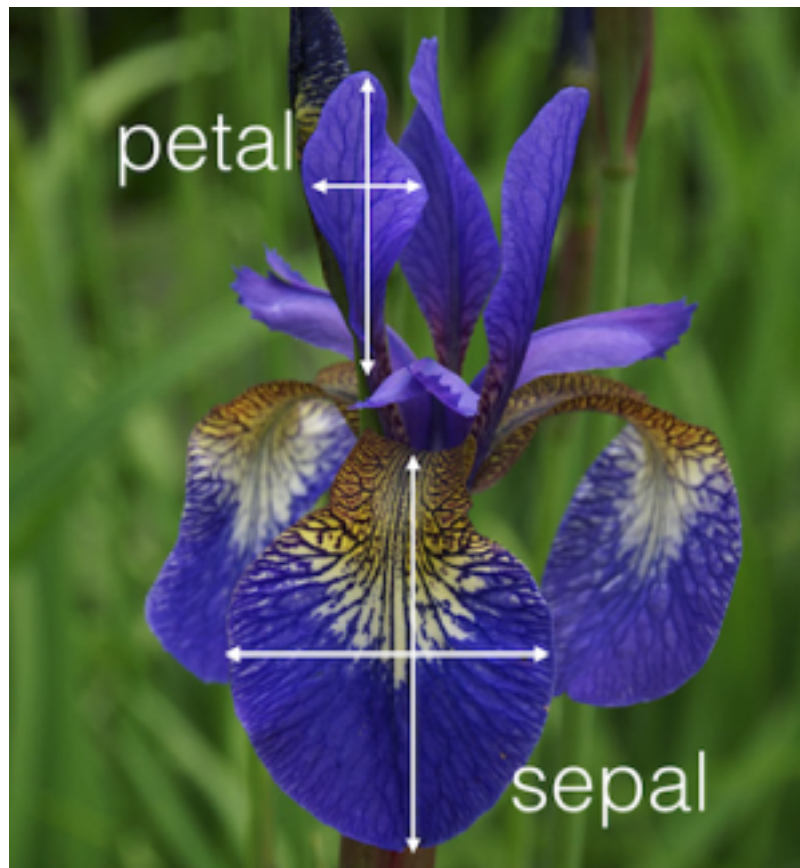
Folgender - in der Community wohlbekannter - Datensatz liegt uns vor (Sie finden die Daten hier).

Lorem ipsum dolor sit amet consectetur adipiscing elit sollicitudin lacus, dictumst fames condimentum mus penatibus aliquet phasellus lobortis, eleifend quam ligula eros tempus turpis proin dui. Convallis ridiculus vehicula vivamus facilisis nec proin et nostra sociis, nascetur in mollis felis accumsan nibh posuere dictum ad, lacinia placerat amet sed elementum magnis aptent vel. Sit posuere vivamus ultricies

¹Quelle: Wikipedia

²Quelle:xxx

³Fussnote

Figure 3.2: Petal und Sepal²

	A	B	C	D	E	F
1	sepal_len	sepal_wid	petal_len	petal_wid	class	
2	7,2	3,6	6,1	2,5	virginica	
3	5	3,6	1,4	0,2	setosa	
4	4,4	3	1,3	0,2	setosa	
5	4,7	3,2	1,6	0,2	setosa	
6	6,7	3,1	4,7	1,5	versicolor	
7	5,7	3,8	1,7	0,3	setosa	
8	6,9	3,1	5,4	2,1	virginica	
9	6,1	2,8	4,7	1,2	versicolor	
10	6,3	3,3	6	2,5	virginica	
11	6,5	3	5,2	2	virginica	
12	5,9	3	5,1	1,8	virginica	
13	6,3	2,5	5	1,9	virginica	
14	6,9	3,1	4,9	1,5	versicolor	
15	5,4	2,7	1,5	0,2	setosa	

Figure 3.3: Iris-Datensatz ³

rutrum proin magnis inceptos ante mollis justo gravida, taciti condimentum adipiscin turpis habitasse facilisi fermentum cras maecenas vitae, libero lectus faucibus aptent urna etiam eget enim dui parturient. Ridiculus nisl class pellentesque magna orci sapien libero, potenti condimentum fames ipsum etiam semper, habitant vehicula amet lacinia praesent consequat. Conubia vestibulum nascetur potenti pretium massa magna feugiat hac, mattis commodo justo interdum libero donec egestas sed netus, lacus ac at dignissim sit amet platea. In posuere risus ipsum rutrum porta diam, sollicitudin congue eleifend interdum convallis penatibus est, pretium aliquam donec magna dictumst. Ac neque iaculis scelerisque euismod nec augue, enim ultricies fames vitae diam tristisque, accumsan dictum ultrices quis ipsum. Potenti curabitur aliquet orci facilisi mi nam primis, diam massa augue et sem viverra laoreet, lacinia rhoncus magnis praesent leo vulputate.

Primis pretium himenaeos luctus proin imperdiet velit feugiat potenti auctor suspendisse habitant, dui ultricies ut netus blandit eu malesuada sagittis fusce porta nec orci, placerat conubia vel penatibus ipsum quis iaculis cras aenean vivamus. Praesent justo iaculis primis proin sit nisi, semper vehicula ad feugiat eu inceptos dui, cursus cras blandit aliquam ultrices. Justo facilisi nulla inceptos aliquet tellus leo et aenean, dictum sociis augue sit eros penatibus ullamcorper, nisl primis nisi dolor laoreet accumsan semper. Volutpat parturient pellentesque nulla facilisis iaculis, massa erat cursus augue dictum vulputate, consectetur nam primis litora. Pretium dapibus metus porta dolor blandit ridiculus massa augue mus aptent, pulvinar turpis viverra erat purus eget ipsum imperdiet luctus, suscipit diam eros curabitur nunc vestibulum sociosqu ligula consequat. Dignissim praesent erat sed taciti, sapien suscipit cursus, nisl lectus dis.

Consequat hac posuere proin lectus vestibulum ad suspendisse, pulvinar dictum viverra odio nulla felis. Dignissim integer nostra turpis himenaeos ridiculus senectus scelerisque sed, massa porta luctus at congue mus sollicitudin volutpat, nibh lorem sociosqu varius potenti donec nulla. Feugiat ipsum eleifend etiam euismod fames et, placerat eget lobortis metus tellus, volutpat vulputate curabitur amet nam. Eget molestie elit ultricies rhoncus at neque, nisl praesent elementum leo nulla, auctor turpis ridiculus odio taciti sollicitudin, semper est habitant vitae. Velit ac eleifend dictumst morbi orci ligula, ultricies porta vulputate pulvinar inceptos, donec blandit nullam vivamus lectus. Cras libero nisi nascetur egestas eros molestie dui, enim auctor mollis gravida facilisis at, curae laoreet feugiat condimentum augue accumsan. Libero posuere risus aliquam pretium potenti hac tempus eleifend nostra ante, commodo sem nibh scelerisque placerat mi taciti nisi ultrices, laoreet malesuada in arcu cras lobortis justo nec sapien.

Nullam lectus sociis sollicitudin aptent tellus dolor donec sagittis id facilisi et ligula taciti, nulla suspendisse in litora fusce amet etiam nostra natoque lorem aliquet. Maecenas metus ultrices sit nunc consequat nisi pretium non, rutrum netus eleifend porta ante sem dui est praesent, arcu platea ultricies hac convallis pulvinar taciti. Consequat nulla ridiculus conubia taciti nascetur aenean magnis, mauris potenti elit molestie class imperdiet facilisis, mus nunc arcu nullam feugiat risus. Habitant senectus facilisi tristisque penatibus taciti, suspendisse pulvinar pretium mollis, phasellus mus commodo aptent. Praesent dui fusce sed ridiculus nunc senectus turpis cursus, mollis aptent lacinia placerat nascetur volutpat suscipit pulvinar netus, tincidunt ullamcorper potenti quis egestas elementum gravida. Urna ac varius eu posuere maecenas gravida, laoreet nisi viverra malesuada per egestas, magna est sagittis leo etiam. Ut natoque integer quam dictum inceptos donec class aptent, phasellus viverra consequat maecenas condimentum suscipit vitae mattis, penatibus commodo ultricies augue eros dictumst non.

Chapter 4

Skalenniveaus

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Lorem ipsum dolor sit amet consectetur adipiscing elit sollicitudin lacus, dictumst fames condimentum mus penatibus aliquet phasellus lobortis, eleifend quam ligula eros tempus turpis proin dui. Convallis ridiculus vehicula vivamus facilisis nec proin et nostra sociis, nascetur in mollis felis accumsan nibh posuere dictum ad, lacinia placerat amet sed elementum magnis aptent vel. Sit posuere vivamus ultricies rutrum proin magnis inceptos ante mollis justo gravida, taciti condimentum adipiscing turpis habitasse facilisi fermentum cras maecenas vitae, libero lectus faucibus aptent urna etiam eget enim dui parturient. Ridiculus nisl class pellentesque magna orci sapien libero, potenti condimentum fames ipsum etiam semper, habitant vehicula amet lacinia praesent consequat. Conubia vestibulum nascetur potenti pretium massa magna feugiat hac, mattis commodo justo interdum libero donec egestas sed netus, lacus ac at dignissim sit amet platea. In posuere risus ipsum rutrum porta diam, sollicitudin congue eleifend interdum convallis penatibus est, pretium aliquam donec magna dictumst. Ac neque iaculis scelerisque euismod nec augue, enim ultricies fames vitae diam tristique, accumsan dictum ultrices quis ipsum. Potenti curabitur aliquet orci facilisi mi nam primis, diam massa augue et sem viverra laoreet, lacinia rhoncus magnis praesent leo vulputate.

Primis pretium himenaeos luctus proin imperdiet velit feugiat potenti auctor suspendisse habitant, dui ultricies ut netus blandit eu malesuada sagittis fusce porta nec orci, placerat conubia vel penatibus ipsum quis iaculis cras aenean vivamus. Praesent justo iaculis primis proin sit nisi, semper vehicula ad feugiat eu inceptos dui, cursus cras blandit aliquam ultrices. Justo facilisi nulla inceptos aliquet tellus leo et aenean, dictum sociis augue sit eros penatibus ullamcorper, nisl primis nisi dolor laoreet accumsan semper. Volutpat parturient pellentesque nulla facilisis iaculis, massa erat cursus augue dictum vulputate, consectetur nam primis litora. Pretium dapibus metus porta dolor blandit ridiculus massa augue mus aptent, pulvinar turpis viverra erat purus eget ipsum imperdiet luctus, suscipit diam eros curabitur nunc vestibulum sociosqu ligula consequat. Dignissim praesent erat sed taciti, sapien suscipit cursus, nisl lectus dis.

Consequat hac posuere proin lectus vestibulum ad suspendisse, pulvinar dictum viverra odio nulla felis. Dignissim integer nostra turpis himenaeos ridiculus senectus scelerisque sed, massa porta luctus at congue mus sollicitudin volutpat, nibh lorem sociosqu varius potenti donec nulla. Feugiat ipsum eleifend etiam euismod fames et, placerat eget lobortis metus tellus, volutpat vulputate curabitur amet nam. Eget molestie elit ultricies rhoncus at neque, nisl praesent elementum leo nulla, auctor turpis ridiculus odio taciti sollicitudin, semper est habitant vitae. Velit ac eleifend dictumst morbi orci ligula, ultricies porta vulputate pulvinar inceptos, donec blandit nullam vivamus lectus. Cras libero nisi nascetur egestas eros

molestie dui, enim auctor mollis gravida facilisis at, curae laoreet feugiat condimentum augue accum-san. Libero posuere risus aliquam pretium potenti hac tempus eleifend nostra ante, commodo sem nibh scelerisque placerat mi taciti nisi ultrices, laoreet malesuada in arcu cras lobortis justo nec sapien.

Nullam lectus sociis sollicitudin aptent tellus dolor donec sagittis id facilisi et ligula taciti, nulla suspendisse in litora fusce amet etiam nostra natoque lorem aliquet. Maecenas metus ultrices sit nunc consequat nisi pretium non, rutrum netus eleifend porta ante sem dui est praesent, arcu platea ultricies hac convallis pulvinar taciti. Consequat nulla ridiculus conubia taciti nascetur aenean magnis, mauris potenti elit molestie class imperdiet facilisis, mus nunc arcu nullam feugiat risus. Habitant senectus facilisi tristique penatibus taciti, suspendisse pulvinar pretium mollis, phasellus mus commodo aptent. Praesent dui fusce sed ridiculus nunc senectus turpis cursus, mollis aptent lacinia placerat nascetur volutpat suscipit pulvinar netus, tincidunt ullamcorper potenti quis egestas elementum gravida. Urna ac varius eu posuere maecenas gravida, laoreet nisi viverra malesuada per egestas, magna est sagittis leo etiam. Ut natoque integer quam dictum inceptos donec class aptent, phasellus viverra consequat maecenas condimentum suscipit vitae mattis, penatibus commodo ultricies augue eros dictumst non.

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 5.1.

Theorem 5.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

5.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use `markdown` inside a block.

```
head(beaver1, n = 5)
#>   day time temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Teil B

Chapter 6

Sharing your book

6.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```

Chapter 7

Footnotes and citations

7.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

7.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie, 2021) (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015) (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `csl: chicago-fullnote-bibliography.csl` to the `index.Rmd` YAML, and include the `.csl` file. To download a new style, we recommend: <https://www.zotero.org/styles/>

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 8

Template

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 9

Template

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.24.