

# Datenanalyse: Verteilung von Fallzahlen und Betten in deutschen Krankenhaus-Stationen (2019)

Patrick Witczak

2025-09-02

## Table of contents

<b>Einleitung [Zielsetzung]</b>	<b>1</b>
<b>Datensatzwahl und Import</b>	<b>2</b>
<b>Datensatzüberblick</b>	<b>4</b>
Variablenbeschreibung . . . . .	4
Verteilung der Bettenzahlen pro Fachabteilung . . . . .	4
Verteilung der Fallzahlen pro Fachabteilung . . . . .	6
<b>Analyse durch Visualisierungen</b>	<b>7</b>
Betten- und Fallzahlen: Korrelationsanalyse . . . . .	7
Top 10 Abteilungen: 80 % der Fallzahlen und entsprechende Bettenzahlen . . . . .	8
<b>Evaluation der Ergebnisse</b>	<b>13</b>
Warum sind die Fallzahlen in Psychatrie-Stationen so niedrig? . . . . .	13
Warum sind die Fallzahlen in der Inneren Medizin so hoch? . . . . .	15
Interpretation . . . . .	16

## Einleitung [Zielsetzung]

In dieser Analyse wird untersucht, wie die Versorgungskapazitäten deutscher Krankenhäuser im Jahr 2019 im Verhältnis zur tatsächlichen Arbeitsbelastung stehen. Dazu werden die Fallzahlen und Bettenzahlen aller Krankenhaus-Stationen gegenübergestellt. Das Jahr 2019 repräsentiert eine typische Krankenhausversorgung vor der COVID-19-Pandemie. Durch

diesen Vergleich lassen sich Unterschiede zwischen vorhandenen Kapazitäten und der Beanspruchung durch Patient:innen aufzeigen. Da solche Unterschiede durch verschiedene Faktoren beeinflusst werden können, werden zusätzlich weitere Variablen herangezogen, um die Analyse zu vertiefen.

## **Datensatzwahl und Import**

Für die Analyse sollten ursprünglich die offiziellen Daten des Statistischen Bundesamtes (Destatis) verwendet werden. Da der direkte Zugriff jedoch kostenpflichtig ist, wurden stattdessen die frei zugänglichen Zahlen aus den Krankenhausreports (Springer Verlag, 2022) herangezogen. Diese beziehen sich auf die offiziellen Destatis-Daten und bilden die Werte für das Jahr 2019 ab. Da der Datensatz mit rund 40 Fachabteilungen eine überschaubare Größe hat, erfolgte der Import manuell über die Funktion `tribble()` aus dem `tidyverse`-Paket. Dadurch konnte der Datensatz direkt in R erstellt und weiterverarbeitet werden:

```

library(tidyverse)
KHR_2019 <- tribble(
  ~Abteilung, ~Abt_Anzahl, ~Abt_Betten,
  ~Nutzungsgrad, ~Fallzahl, ~Avg_Verweildauer,
  "Innere Medizin", 1047, 111481, 78.1, 5889078, 5.4,
  "Geriatrie", 322, 18101, 86.9, 374462, 15.3,
  "Kardiologie", 206, 14800, 83.6, 955424, 4.7,
  "Nephrologie", 62, 2186, 83.5, 96053, 6.9,
  "Hämatologie und internistische Onkologie", 102, 5096, 79.9, 201198, 7.4,
  "Endokrinologie", 18, 501, 83.3, 23432, 6.5,
  "Gastroenterologie", 126, 7203, 82.5, 418259, 5.2,
  "Pneumologie", 64, 3689, 79.4, 178007, 6.0,
  "Rheumatologie", 33, 1096, 72.1, 38126, 7.6,
  "Pädiatrie", 339, 15447, 63.2, 926516, 3.8,
  "Kinderkardiologie", 23, 588, 65.0, 19087, 7.3,
  "Neonatologie", 99, 2097, 72.3, 56492, 9.8,
  "Kinderchirurgie", 83, 1592, 63.8, 124763, 3.0,
  "Lungen- und Bronchialheilkunde", 18, 1775, 69.5, 62763, 7.2,
  "Allgemeine Chirurgie", 1068, 67902, 69.8, 3195674, 5.4,
  "Unfallchirurgie", 308, 17303, 78.1, 840038, 5.9,
  "Neurochirurgie", 178, 6642, 77.6, 256956, 7.3,
  "Gefäßchirurgie", 186, 5529, 72.3, 189430, 7.7,
  "Plastische Chirurgie", 132, 1936, 69.2, 89349, 5.5,
  "Thoraxchirurgie", 56, 1546, 67.8, 50570, 7.6,
  "Herzchirurgie", 71, 4614, 77.5, 152524, 8.6,
  "Urologie", 500, 14036, 74.4, 886904, 4.3,
  "Orthopädie", 435, 23078, 68.5, 904191, 6.4,
  "Frauenheilkunde und Geburtshilfe", 759, 25039, 60.8, 1548418, 3.6,
  "Geburtshilfe", 86, 2205, 78.3, 180552, 3.5,
  "Hals-, Nasen-, Ohrenheilkunde", 592, 8672, 60.5, 562945, 3.4,
  "Augenheilkunde", 264, 4219, 65.4, 360377, 2.8,
  "Neurologie", 455, 26716, 81.2, 1092503, 7.2,
  "Allgemeine Psychiatrie", 394, 57269, 94.0, 769076, 24.7,
  "Kinder- und Jugendpsychiatrie", 144, 6696, 88.7, 59578, 36.4,
  "Psychosomatik/Psychotherapie", 275, 12394, 90.1, 93160, 43.7,
  "Nuklearmedizin", 96, 724, 44.9, 36061, 3.3,
  "Strahlenheilkunde", 145, 2489, 68.8, 70938, 8.8,
  "Dermatologie", 109, 4575, 77.3, 240506, 5.4,
  "Zahn- und Kieferheilkunde, Mund- und Kieferchirurgie",
  183, 2041, 65.4, 116377, 4.2,
  "Intensivmedizin", 260, 6568, 77.0, 474544, 3.9,
  "Sonstige Fachabteilung", 305, 6481, 73.5, 269292, 6.5)

```

## Datensatzüberblick

### Variablenbeschreibung

Der Datensatz umfasst die folgenden Variablen:

#### Primäre Variablen für die Analyse

**Abt\_Betten:** Gesamtzahl der Betten je Fachabteilung

**Fallzahl:** Gesamtzahl der Fälle je Fachabteilung

#### Zusätzliche Variablen für vertiefende Analysen

**Abt\_Anzahl:** Anzahl der Einrichtungen mit der jeweiligen Fachabteilung

**Nutzungsgrad:** durchschnittlicher Belegungsgrad der Betten (in %)

**Avg\_Verweildauer:** durchschnittliche Verweildauer pro Fall (in Tagen)/li>

```
head(KHR_2019)
```

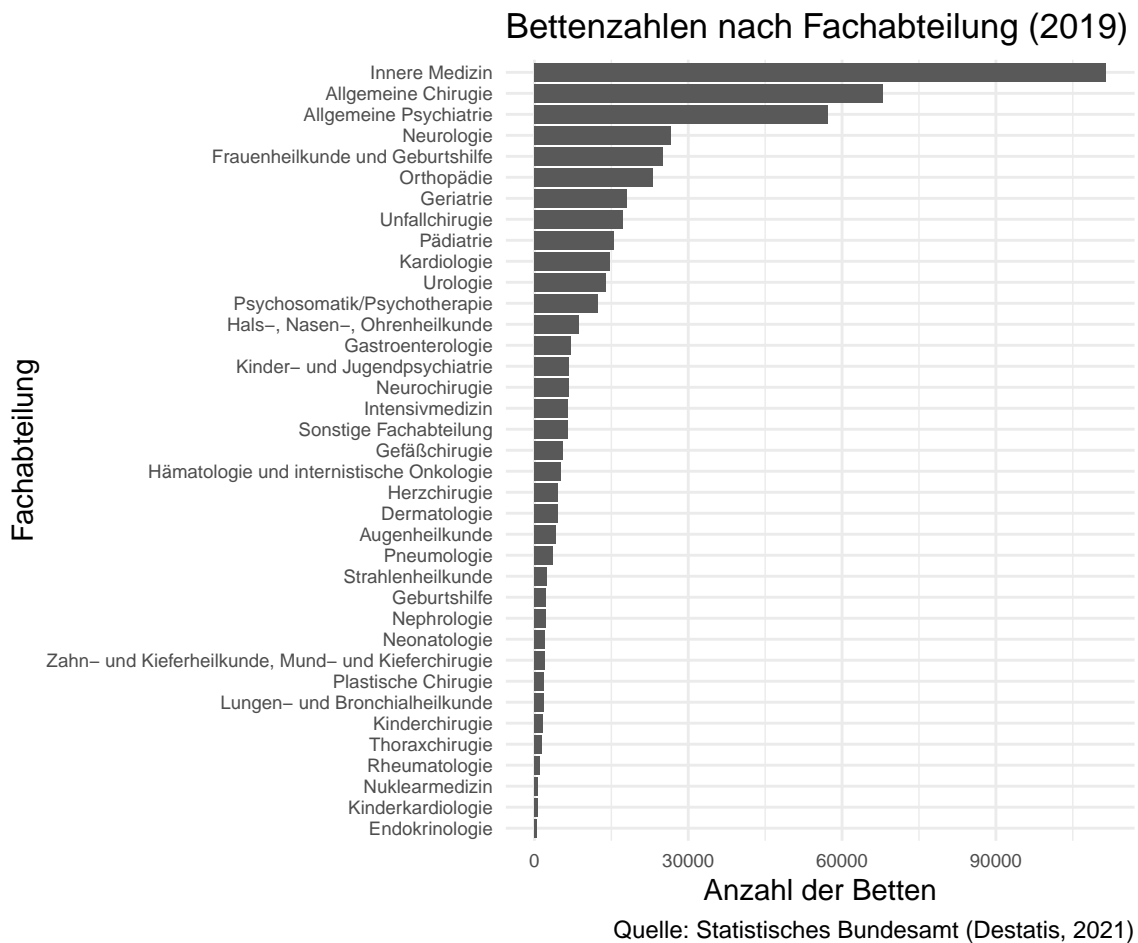
```
# A tibble: 6 x 6
```

Abteilung <chr>	Abt_Anzahl <dbl>	Abt_Betten <dbl>	Nutzungsgrad <dbl>	Fallzahl <dbl>	Avg_Verweildauer <dbl>
1 Innere Medizin	1047	111481	78.1	5889078	5.4
2 Geriatrie	322	18101	86.9	374462	15.3
3 Kardiologie	206	14800	83.6	955424	4.7
4 Nephrologie	62	2186	83.5	96053	6.9
5 Hämatologie und ~	102	5096	79.9	201198	7.4
6 Endokrinologie	18	501	83.3	23432	6.5

### Verteilung der Bettenzahlen pro Fachabteilung

Die Gegenüberstellung der Bettenzahlen zeigt eine starke Konzentration: Weniger als die Hälfte der 37 Fachabteilungen vereint schätzungsweise mehr als 80 % der gesamten Krankenhausbetten. Dies verdeutlicht die zentrale Bedeutung großer Basisfächer wie Innere Medizin oder Allgemeine Chirurgie, während hochspezialisierte Fachrichtungen naturgemäß über deutlich kleinere Bettenkapazitäten verfügen.

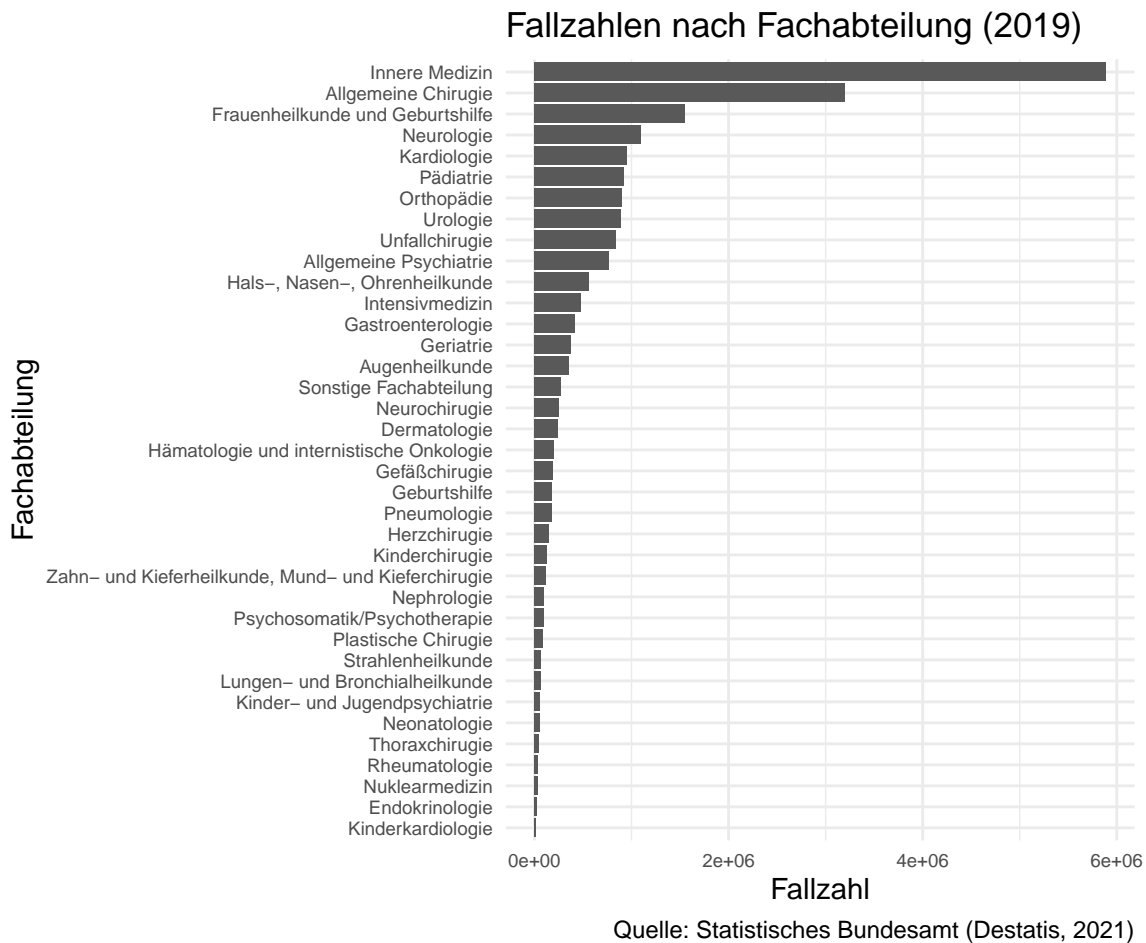
```
KHR_2019 |>
  mutate(Abteilung = fct_reorder(Abteilung, Abt_Betten)) |>
  ggplot(aes(y = Abteilung, x = Abt_Betten)) +
  geom_col() +
  labs(
    title = "Bettenzahlen nach Fachabteilung (2019)",
    x = "Anzahl der Betten",
    y = "Fachabteilung",
    caption = "Quelle: Statistisches Bundesamt (Destatis, 2021)") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 7)) +
  theme(axis.text.x = element_text(size = 7))
```



## Verteilung der Fallzahlen pro Fachabteilung

Die Verteilung der Fallzahlen zeigt ein ähnliches Muster wie die Bettenzahlen: Ein Großteil der Fälle konzentriert sich auf wenige große Fachabteilungen. Auf den ersten Blick erscheinen die beiden Variablen eng miteinander verknüpft, wenngleich es in einzelnen Abteilungen Abweichungen gibt. Diese Beziehung wird im nächsten Analyseschritt anhand eines Scatterplots überprüft.

```
KHR_2019 |>
  mutate(Abteilung = fct_reorder(Abteilung, Fallzahl)) |>
  ggplot(aes(y = Abteilung, x = Fallzahl)) +
  geom_col() +
  labs(
    title = "Fallzahlen nach Fachabteilung (2019)",
    x = "Fallzahl",
    y = "Fachabteilung",
    caption = "Quelle: Statistisches Bundesamt (Destatis, 2021)") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 7)) +
  theme(axis.text.x = element_text(size = 7))
```



## Analyse durch Visualisierungen

### Betten- und Fallzahlen: Korrelationsanalyse

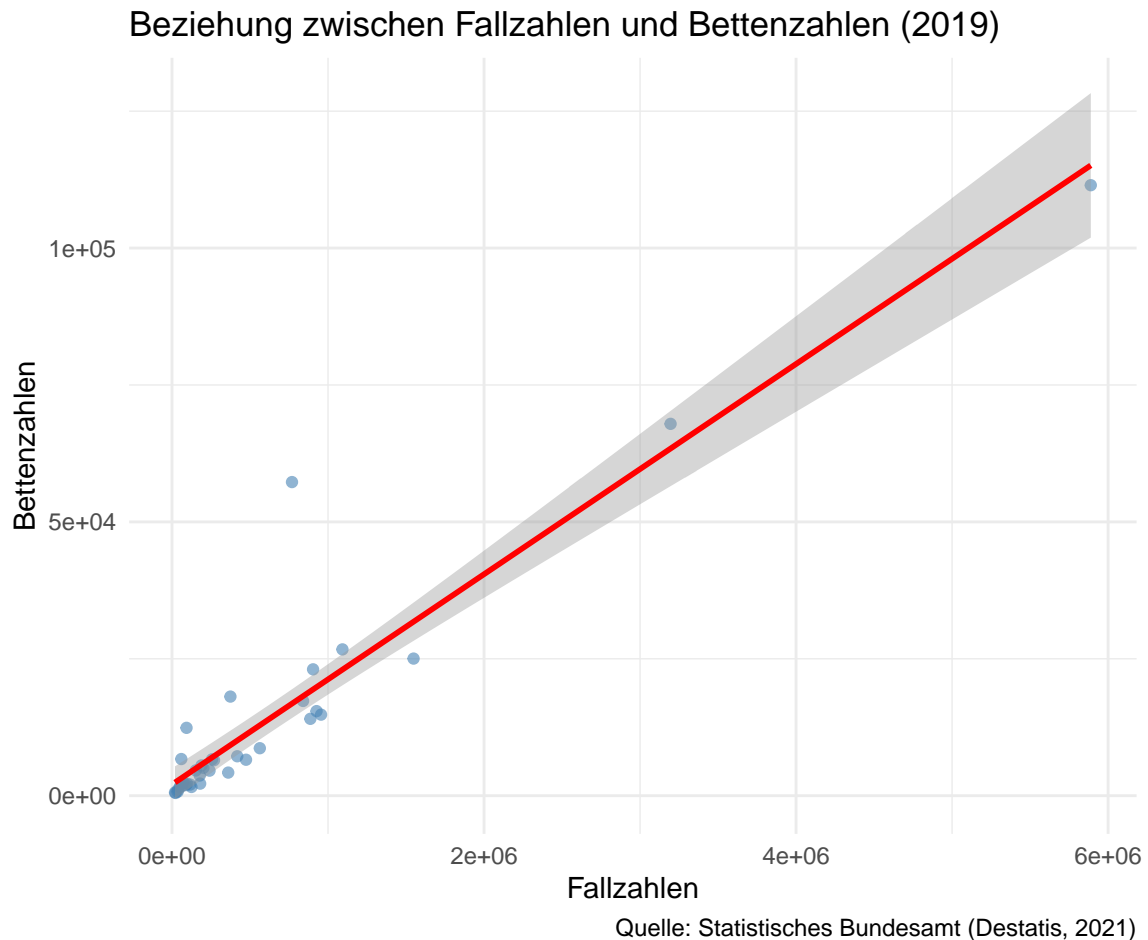
Die Analyse deutet auf eine positive lineare Beziehung zwischen Fallzahlen und Bettenzahlen hin: Abteilungen mit höheren Fallzahlen verfügen tendenziell über mehr Betten. Trotz dieser allgemeinen Linearität zeigen einzelne Abteilungen Abweichungen von der Trendlinie. Aufgrund der geringen Anzahl von 37 Beobachtungen sollten Interpretationen vorsichtig erfolgen. Weiterführende Analysen können mögliche Ursachen dieser Abweichungen untersuchen.

```
ggplot(KHR_2019, aes(x = Fallzahl, y = Abt_Betten)) +
  geom_point(alpha = 0.6, color = "steelblue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
```

```

title = "Beziehung zwischen Fallzahlen und Bettenzahlen (2019)",
x = "Fallzahlen",
y = "Bettenzahlen",
caption = "Quelle: Statistisches Bundesamt (Destatis, 2021)" +
theme_minimal()

```



## Top 10 Abteilungen: 80 % der Fallzahlen und entsprechende Bettenzahlen

### Problemstellungen

Ziel der Analyse war es, eine direkte Gegenüberstellung von Fallzahlen und Bettenzahlen deutscher Krankenhaus-Stationen zu erstellen. Dabei traten folgende Herausforderungen auf:



Die Daten enthalten viele hochspezialisierte Stationen mit niedrigen Fallzahlen, die die Visualisierung unübersichtlich machen.

Die ursprüngliche Struktur des Datensatzes ist für einen direkten Vergleich der beiden Kennzahlen nicht geeignet.

Fallzahlen und Bettenzahlen liegen in sehr unterschiedlichen Größenordnungen, was eine Normalisierung notwendig macht.

Diese Probleme führten dazu, dass **Pareto-Logik** und eine **Umstrukturierung des Datensatzes** angewendet werden mussten.

## Pareto Logik

Um die Anzahl der dargestellten Stationen zu reduzieren und dennoch repräsentative Ergebnisse zu erhalten, wurde die Pareto-Logik auf die Fallzahlen angewendet. Es wurden nur die größten Stationen ausgewählt, die zusammen 80 % der Fallzahlen abdecken. So reduzierte sich die Stationsanzahl von 37 auf 10, wodurch die Visualisierung übersichtlich bleibt und die wichtigsten Abteilungen hervorgehoben werden.

```
KHR_2019 |>
  arrange(desc(Fallzahl)) |>
  mutate(rel_cumsum_Fallzahl = cumsum(Fallzahl) / sum(Fallzahl)) |>
  # Top 80% der Abteilungen
  filter(rel_cumsum_Fallzahl <= 0.8) |>
  select(Abteilung, Fallzahl, rel_cumsum_Fallzahl)
```

# A tibble: 10 x 3

Abteilung	Fallzahl	rel_cumsum_Fallzahl
<chr>	<dbl>	<dbl>
1 Innere Medizin	5889078	0.270
2 Allgemeine Chirurgie	3195674	0.417
3 Frauenheilkunde und Geburtshilfe	1548418	0.488
4 Neurologie	1092503	0.538
5 Kardiologie	955424	0.582
6 Pädiatrie	926516	0.624
7 Orthopädie	904191	0.666
8 Urologie	886904	0.706
9 Unfallchirurgie	840038	0.745
10 Allgemeine Psychiatrie	769076	0.780

## 02 Änderung der Datensatzstruktur

Für die Visualisierung ist es unpraktisch, wenn Bettenzahlen und Fallzahlen in getrennten Spalten stehen. Daher wurden die beiden numerischen Variablen in eine numerische Variable (Wert) und eine kategoriale Variable (Kennzahl) transformiert:

Kennzahl enthält die Bezeichnung der ursprünglichen Variablen („Abt\_Betten“ oder „Fallzahl“).

Wert enthält die zugehörigen Zahlen.

```
KHR_2019 |>
# 2 Hauptvariabel-Titel werden zu Spalten im Tibble
pivot_longer(cols = c(Abt_Betten, Fallzahl),
              names_to = "Kennzahl",
              values_to = "Wert") |>
select(Abteilung, Kennzahl, Wert) |>
head()
```

```
# A tibble: 6 x 3
  Abteilung      Kennzahl      Wert
  <chr>         <chr>      <dbl>
1 Innere Medizin Abt_Betten  111481
2 Innere Medizin Fallzahl    5889078
3 Geriatrie      Abt_Betten   18101
4 Geriatrie      Fallzahl    374462
5 Kardiologie    Abt_Betten   14800
6 Kardiologie    Fallzahl    955424
```

Diese Transformation ermöglicht es, beide Kennzahlen in einem einzigen Plot darzustellen.

## 03 Relativierung der Zahlen

Da die Bettenzahlen fünfstellige Werte widerspiegeln, die Fallzahlen jedoch siebenstellige, ist ein direkter Vergleich nicht sinnvoll. Zur Vergleichbarkeit wurden die Werte über die Funktion `geom_col(position = "fill")` auf 100 % skaliert, sodass die Balken die relativen Anteile der Abteilungen darstellen.

## Die Analyse

In Form von **zwei gestapelten 100%-Balkendiagrammen** können die Verhältnisse von Fallzahlen und Bettenzahlen direkt verglichen werden. Diese Darstellung verdeutlicht, welche Abteilungen über- oder unterproportional mit Betten ausgestattet sind und liefert erste Hinweise auf mögliche Besonderheiten in der Kapazitätsplanung.

Während die meisten Abteilungen nahezu identische relative Anteile in beiden Kennzahlen zeigen, fallen die folgenden Ausreißer besonders auf:

**Allgemeine Psychiatrie**, bei der die relative Bettenzahl etwa dreimal so hoch ist wie die relative Fallzahl.

**Innere Medizin**, bei der die relativen Fallzahlen rund 5 % höher liegen als die Bettenzahlen.

```
### Laden der Viridis-Bibliothek für das Farbschema
library(viridis)

### Rechnung
KHR_2019_long <- KHR_2019 |>
  # Sortiere Stationen nach Fallzahlen statt alphabetisch
  mutate(Abteilung = fct_reorder(Abteilung, Fallzahl, .desc = TRUE)) |>
  # Für bessere Lesbarkeit in der Visualisierung
  rename(Bettenzahl = Abt_Betten) |>
  # Top 80% der Abteilungen
  arrange(desc(Fallzahl)) |>
  mutate(rel_cumsum_Fallzahl = cumsum(Fallzahl) / sum(Fallzahl)) |>
  filter(rel_cumsum_Fallzahl <= 0.8) |>
  # 2 Hauptvariabel-Titel werden zu Spalten im Tibble
  pivot_longer(cols = c(Bettenzahl, Fallzahl),
               names_to = "Kennzahl",
               values_to = "Wert") |>
  # Bug-Fix: paste0() funktionierte nicht - Rechnung wurde ausgegliedert
  group_by(Kennzahl) |>
  mutate(Anteil = Wert / sum(Wert))

### Visualisierung
ggplot(data = KHR_2019_long, aes(
  x = Kennzahl, y = Wert, fill = Abteilung)) +
  # Alle Variablen skalieren unabhängig bis 100% dank position = "fill"
  geom_col(position = "fill") +
  # Fallzahlen separiert (skalieren unterschiedlich)
  geom_text(data = filter(KHR_2019_long, Kennzahl == "Fallzahl"),
            # Rechnung wurde ausgegliedert - Bug-Fix
```

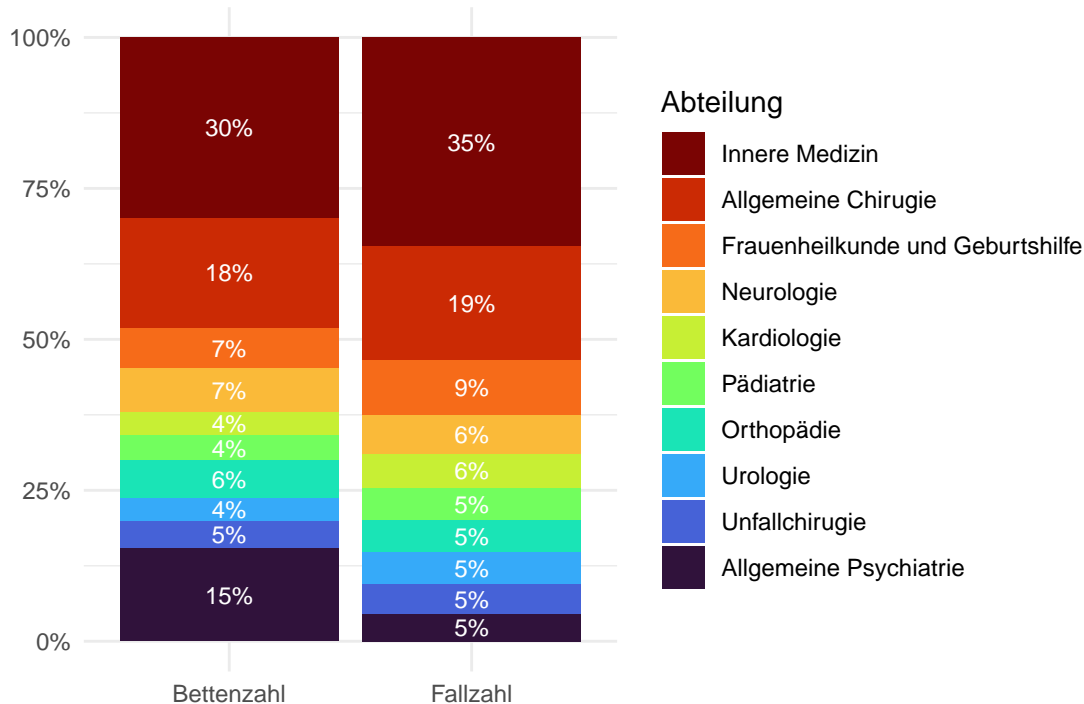
```

    aes(label = paste0(round(Anteil * 100, 0), "%")),
    # Notwendig, damit Zahl korrekt Diagramm überlappt
    position = position_fill(vjust = 0.5),
    color = "white",
    size = 3) +
# Bettenzahlen separiert (skalieren unterschiedlich)
geom_text(data = filter(KHR_2019_long, Kennzahl == "Bettenzahl"),
    # Rechnung wurde ausgegliedert - Bug-Fix
    aes(label = paste0(round(Anteil * 100, 0), "%")),
    # Notwendig, damit Zahl korrekt Diagramm überlappt
    position = position_fill(vjust = 0.5),
    color = "white",
    size = 3) +
scale_y_continuous(labels = scales::percent) +
theme_minimal() +
# Farbschema [viridis Bibliothek]
scale_fill_viridis_d(option = "turbo", direction = -1) +
labs(x = "", y = "",
    title = "Verteilung von Fallzahlen und Betten
in deutschen Krankenhaus-Stationen (2019)",
    subtitle = "Top 10 Abteilungen,
die zusammen 80% aller Fallzahlen abdecken",
    caption = "Quelle: Statistisches Bundesamt (Destatis, 2021)")

```

## Verteilung von Fallzahlen und Betten in deutschen Krankenhaus-Stationen (2019)

Top 10 Abteilungen,  
die zusammen 80% aller Fallzahlen abdecken



Quelle: Statistisches Bundesamt (Destatis, 2021)

## Evaluation der Ergebnisse

### Warum sind die Fallzahlen in Psychiatrie-Stationen so niedrig?

Ein Blick auf die durchschnittliche Verweildauer der Patienten liefert eine Erklärung: In Psychiatrien verbleiben Patient:innen im Durchschnitt fast 25 Tage, während auf den meisten anderen Stationen der Durchschnitt bei etwa 5 Tagen liegt. Die längere Belegung führt zu einer niedrigeren Patientenrotation, was die vergleichsweise niedrigen Fallzahlen erklärt.

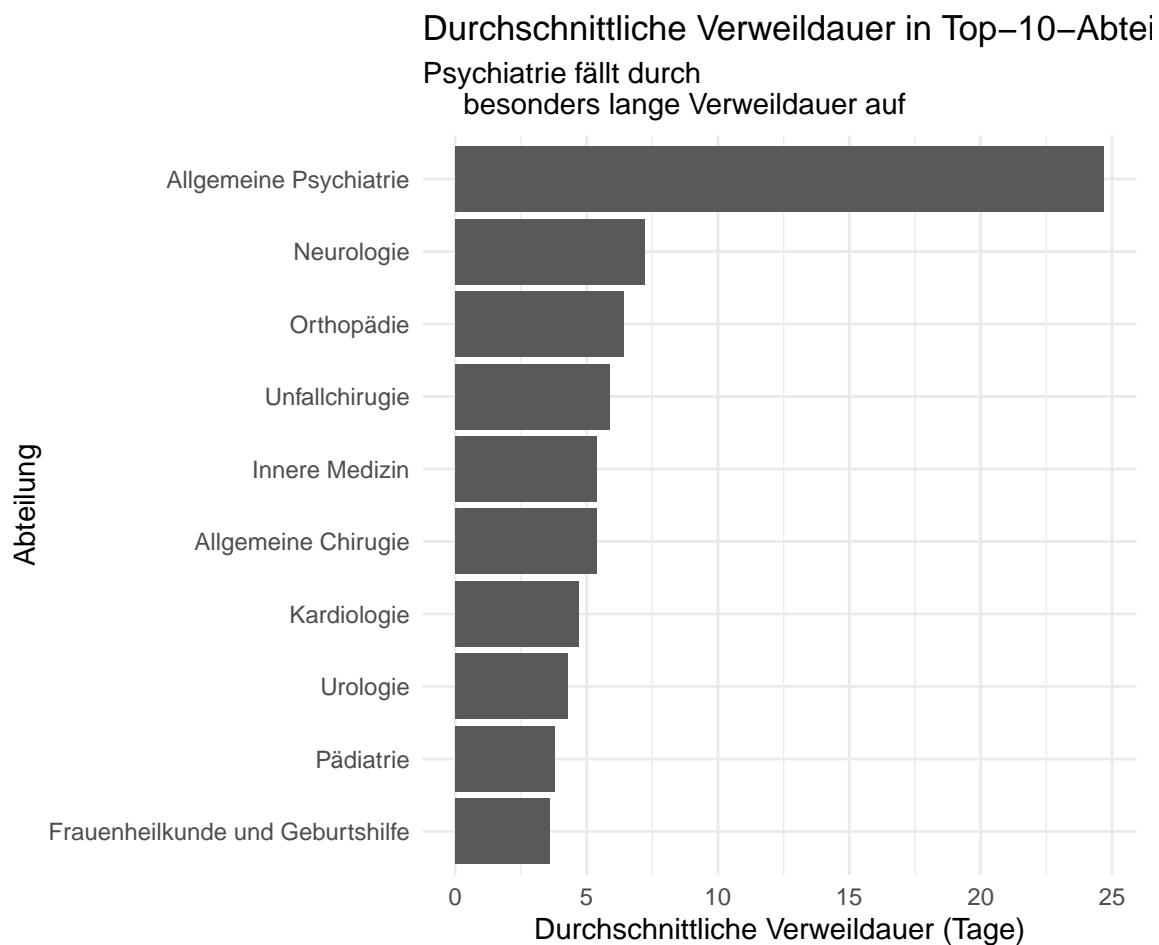
Der folgende Plot zeigt die durchschnittliche Verweildauer der Top-10-Abteilungen und verdeutlicht die besonderen Werte der Psychiatrie-Stationen:

```
KHR_2019 |>
  arrange(desc(Fallzahl)) |>
```

```

mutate(rel_cumsum_Fallzahl = cumsum(Fallzahl) / sum(Fallzahl)) |>
filter(rel_cumsum_Fallzahl <= 0.8) |>
mutate(Abteilung = fct_reorder(Abteilung, Avg_Verweildauer)) |>
ggplot(aes(y = Abteilung, x = Avg_Verweildauer)) +
geom_col() +
labs(
  x = "Durchschnittliche Verweildauer (Tage)",
  y = "Abteilung",
  title = "Durchschnittliche Verweildauer in Top-10-Abteilungen",
  subtitle = "Psychiatrie fällt durch  
besonders lange Verweildauer auf") +
theme_minimal()

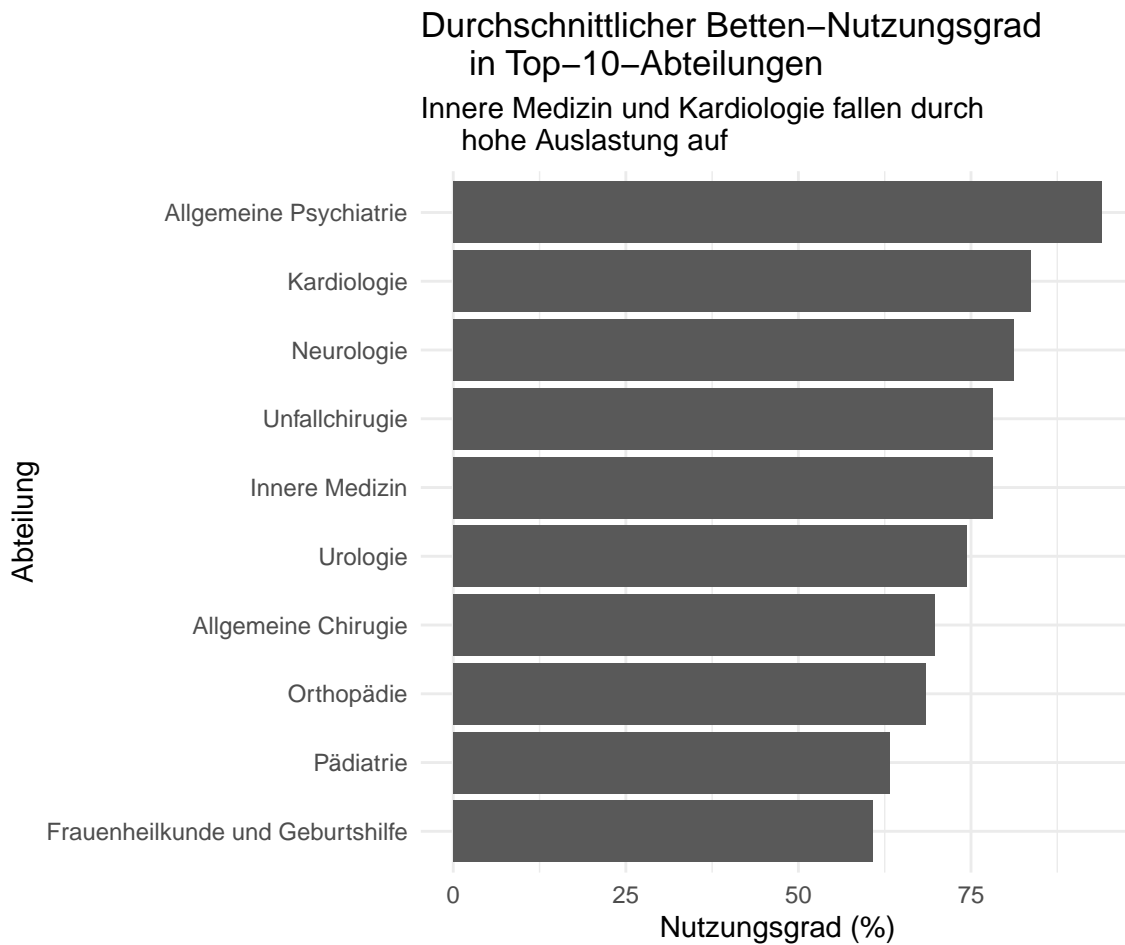
```



## Warum sind die Fallzahlen in der Inneren Medizin so hoch?

Bei Miteinbezug aller verfügbaren Variablen konnten keine klaren Ursachen für das höhere Verhältnis von Fallzahlen zu Betten festgestellt werden. Interessanterweise zeigt die Kardiologie ein ähnliches, weniger deutliches Verhältnis, was vermutlich durch die inhaltlichen Überschneidungen der beiden Fachbereiche bedingt ist. Es ist bekannt, dass Pflegekräfte in der Inneren Medizin und Kardiologie unter vergleichsweise hohem Arbeitsdruck stehen. Diese Annahme lässt sich mit den vorliegenden Daten nicht statistisch untermauern, liefert jedoch einen plausiblen Kontext. Der folgende Plot zeigt den durchschnittlichen Nutzungsgrad der Top-10-Abteilungen:

```
KHR_2019 |>
  arrange(desc(Fallzahl)) |>
  mutate(rel_cumsum_Fallzahl = cumsum(Fallzahl) / sum(Fallzahl)) |>
  filter(rel_cumsum_Fallzahl <= 0.8) |>
  mutate(Abteilung = fct_reorder(Abteilung, Nutzungsgrad)) |>
  ggplot(aes(y = Abteilung, x = Nutzungsgrad)) +
  geom_col() +
  labs(
    x = "Nutzungsgrad (%)",
    y = "Abteilung",
    title = "Durchschnittlicher Betten-Nutzungsgrad
in Top-10-Abteilungen",
    subtitle = "Innere Medizin und Kardiologie fallen durch
hohe Auslastung auf") +
  theme_minimal()
```



### Interpretation

Die hohe Auslastung in der Inneren Medizin könnte ein Hinweis auf einen höheren Bedarf an Betten oder Personal in diesem Bereich sein. Konkrete Handlungsempfehlungen lassen sich aus diesen Daten jedoch nicht ableiten, können aber in einer weiterführenden Planung diskutiert werden.