

# 数据分析：德国医院病房病例数和床位分布（2019 年）

Patrick Witczak

2025-09-02

## Table of contents

引言 [目标]	2
数据集选择与导入	2
数据集概览	4
变量描述 . . . . .	4
各科室床位分布 . . . . .	4
各科室病例数分布 . . . . .	6
通过可视化进行分析	7
床位数量与病例数：相关性分析 . . . . .	7
前 10 个科室：80% 的病例量及其对应的床位数量 . . . . .	8
结果评估	13
为什么精神科病房的病例数如此之低? . . . . .	13
为什么内科的病例数如此之高? . . . . .	15
解读 . . . . .	16

## 引言 [目标]

本分析旨在考察 2019 年德国医院的护理能力与实际工作量之间的关系。为此，我们将比较所有病房的病例数量和床位数量。2019 年代表了新冠疫情爆发前的典型医院护理状况。这一比较揭示了可用容量与患者需求之间的差异。由于这种差异可能受多种因素影响，因此我们引入了其他变量来进一步深化分析。

## 数据集选择与导入

本分析最初计划使用联邦统计局 (Destatis) 的官方数据。但由于直接访问需要付费，因此我们改用了医院报告中可免费访问的数据 (Springer Verlag, 2022)。这些数据指的是 Destatis 的官方数据，代表的是 2019 年的数据。由于数据集规模适中，包含约 40 个科室，因此我们使用 tidyverse 包中的 tribble() 函数手动导入。这样就可以直接在 R 中创建并进一步处理数据集：

```
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)

## 数据集概览

### 变量描述

该数据集包含以下变量：

#### 主要分析变量

科室床位：每个科室的床位总数

病例数：每个科室的病例总数

#### 其他深入分析变量

科室编号：相应科室的医疗机构数量

利用率：平均床位占用率（单位：%）

平均住院时长：每例病例的平均住院时长（单位：天）

```
head(KHR_2019)
```

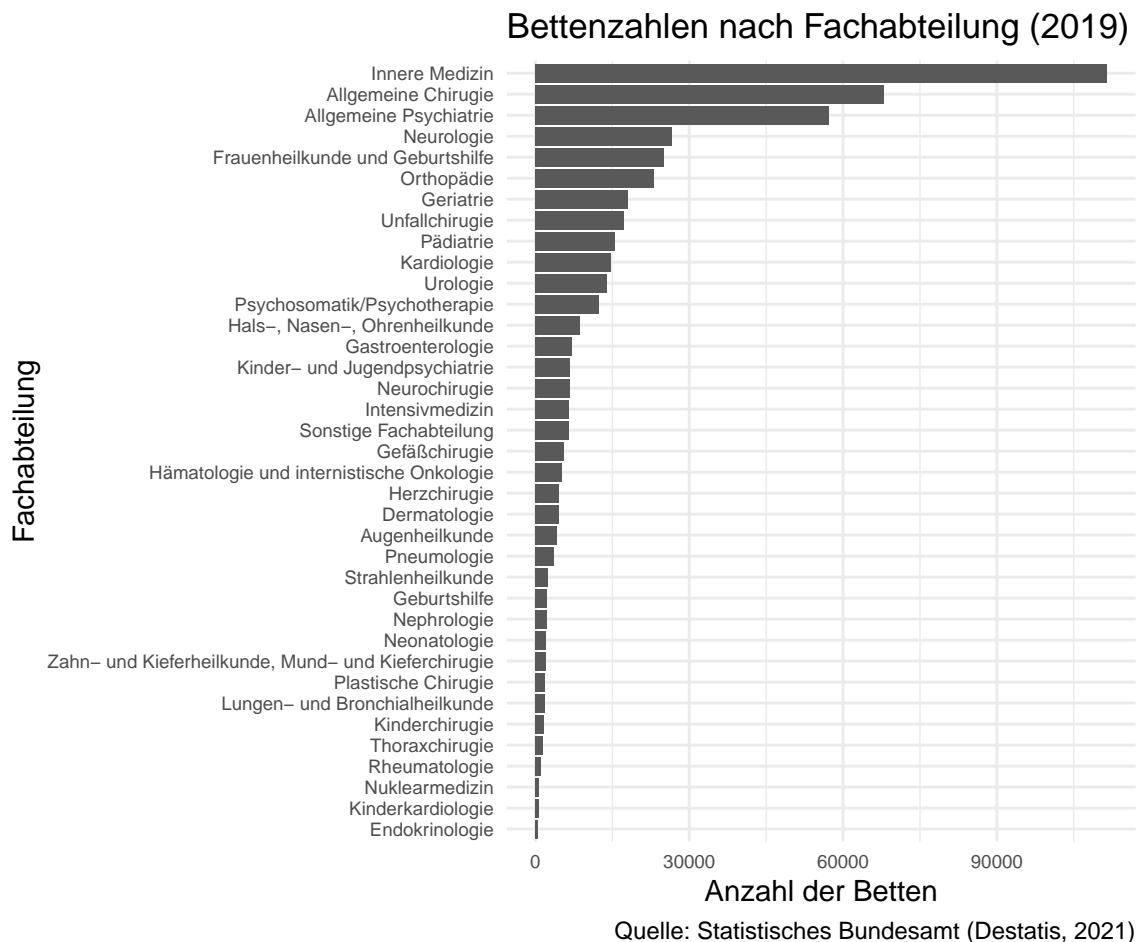
```
# A tibble: 6 x 6
```

Abteilung	Abt_Anzahl	Abt_Betten	Nutzungsgrad	Fallzahl	Avg_Verweildauer
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<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

### 各科室床位分布

床位数量对比显示，医院床位高度集中：37 个科室中，不到一半的科室占据了医院总床位的约 80% 以上。这凸显了内科或普外科等主要基础专科的重要性，而高度专科化的专科床位容量自然要小得多。

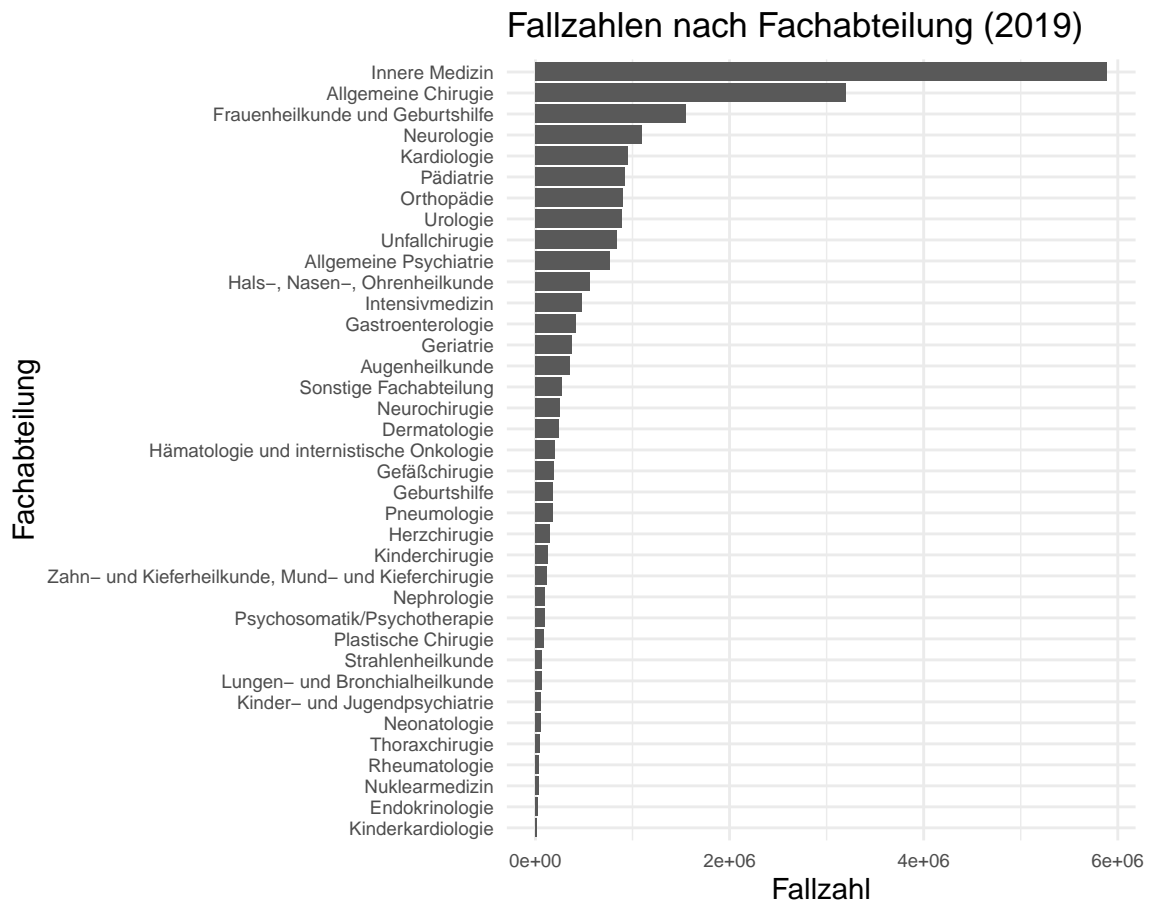
```
KHR_2019 |>
  mutate(Abteilung = fct_reorder(Abteilung, Abt_Betten)) |>
  ggplot(aes(y = Abteilung, x = Abt_Betten)) +
  geom_col() +
  labs(
    title = "Bettenzahlen nach Fachabteilung (2019)",
    x = "Anzahl der Betten",
    y = "Fachabteilung",
    caption = "Quelle: Statistisches Bundesamt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



## 各科室病例数分布

病例数分布与床位数量呈现相似趋势：大量病例集中在少数几个较大的科室。乍一看，这两个变量似乎密切相关，但个别科室之间存在偏差。下一步分析将使用散点图来检验这种关系。

```
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))
```



Quelle: Statistisches Bundesamt (Destatis, 2021)

## 通过可视化进行分析

### 床位数量与病例数：相关性分析

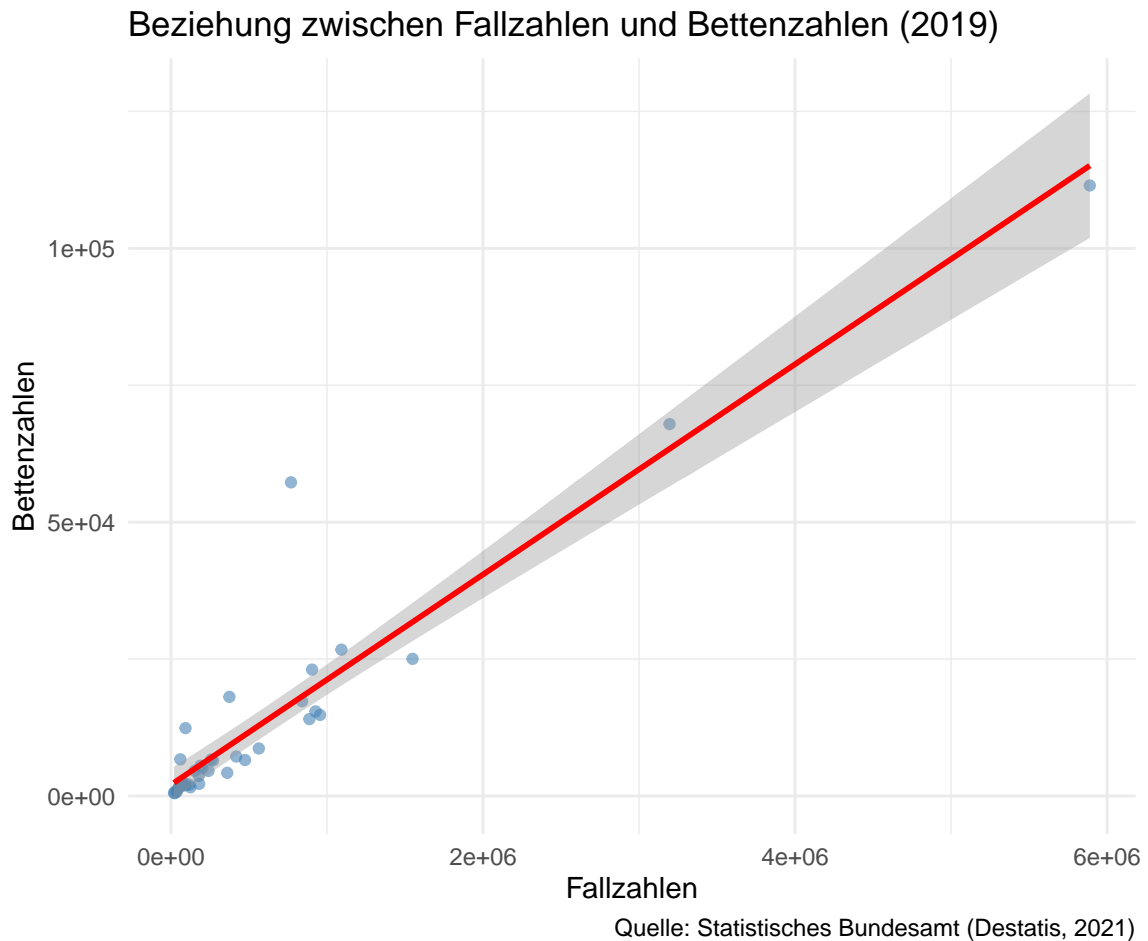
分析表明，病例数与床位数量呈正线性关系：病例数较高的科室往往拥有更多床位。尽管总体呈线性关系，但个别科室仍存在与趋势线的偏差。由于观测数据较少，仅有 37 个，因此解读时应谨慎。进一步分析可以探究造成这些偏差的可能原因。

```
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()

```



前 10 个科室：80% 的病例量及其对应的床位数量

问题

本分析的目标是直接比较德国医院病房的病例量和床位数量。出现了以下挑战：

数据包含许多高度专业化的病房，但病例量较低，这使得可视化效果令人困惑。



数据集的原始结构不适合直接比较这两个关键数据。

病例量和床位数量在数量上差异很大，因此有必要进行标准化。

这些问题促使我们应用帕累托逻辑并重构数据集。

#### 帕累托逻辑

为了减少显示的病房数量，同时仍能获得具有代表性的结果，我们对案件量应用了帕累托逻辑。我们仅选择了最大的病房，这些病房合计占案件量的 80%。这使得病房数量从 37 个减少到 10 个，从而保持了可视化的清晰度，并突出了最重要的科室。

```
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 数据集结构变更

对于可视化而言，将床位数量和病例数量分别放在不同的列中是不切实际的。因此，将两个数值变量转换为一个数值变量（值）和一个分类变量（关键值）：

“关键值”包含原始变量的名称（“Dept\_Beds”或“Case count”）。

“值”包含相应的数值。

```
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
```

这种转换使得两个关键数据可以在一个图中显示。

## 03 数值相对化

由于床位数是五位数，而病例数是七位数，因此直接比较意义不大。为了便于比较，我们使用 `geom_col(position = "fill")` 函数将数值缩放到 100%，以便条形图代表各科室的相对比例。

## 分析

病例数和床位数之间的比率可以通过两个堆叠的 **100% 条形图** 直接比较。这种表示方式可以明确哪些科室的床位配置不均衡或不均衡，并为容量规划中可能出现的特殊情况提供初步指示。

虽然大多数科室在两项指标中的相对份额几乎相同，但以下异常值尤为引人注目：

普通精神病学，其相对床位数约为相对病例数的三倍。

内科，其相对病例数约比床位数高出 5%。

```
### 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") +
```

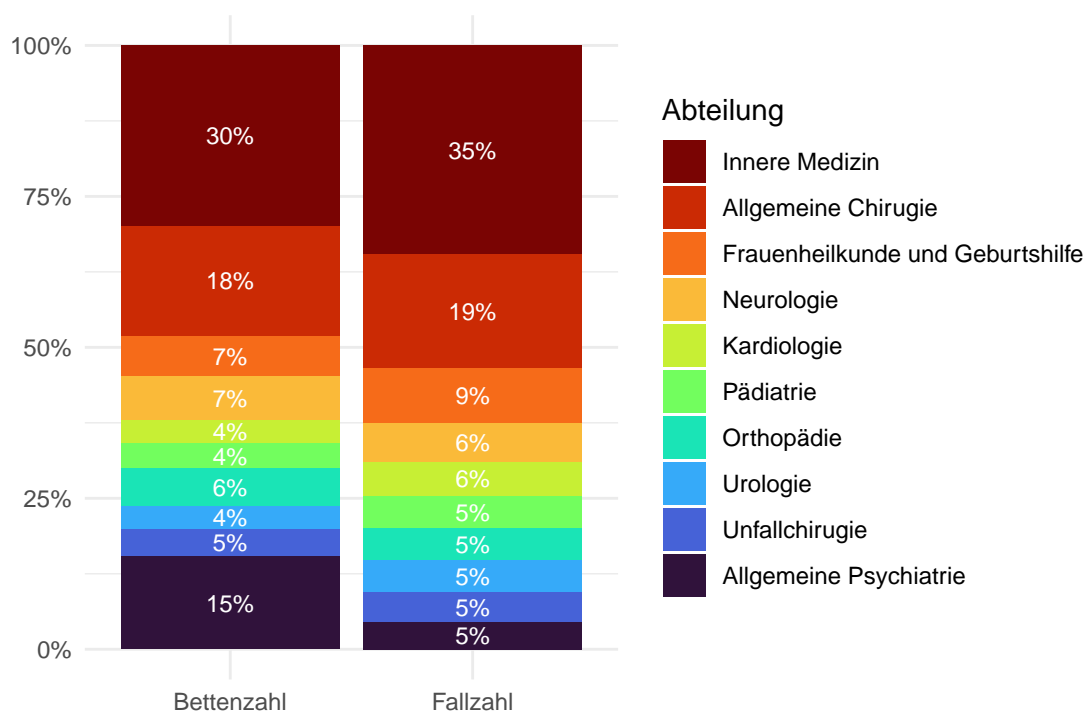
```

# 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)

## 结果评估

为什么精神科病房的病例数如此之低？

查看平均住院日数可以解释这一点：精神科病房的患者平均住院日数接近 25 天，而大多数其他病房的平均住院日数约为 5 天。住院时间越长，患者轮换率就越低，这解释了病例数相对较低的原因。

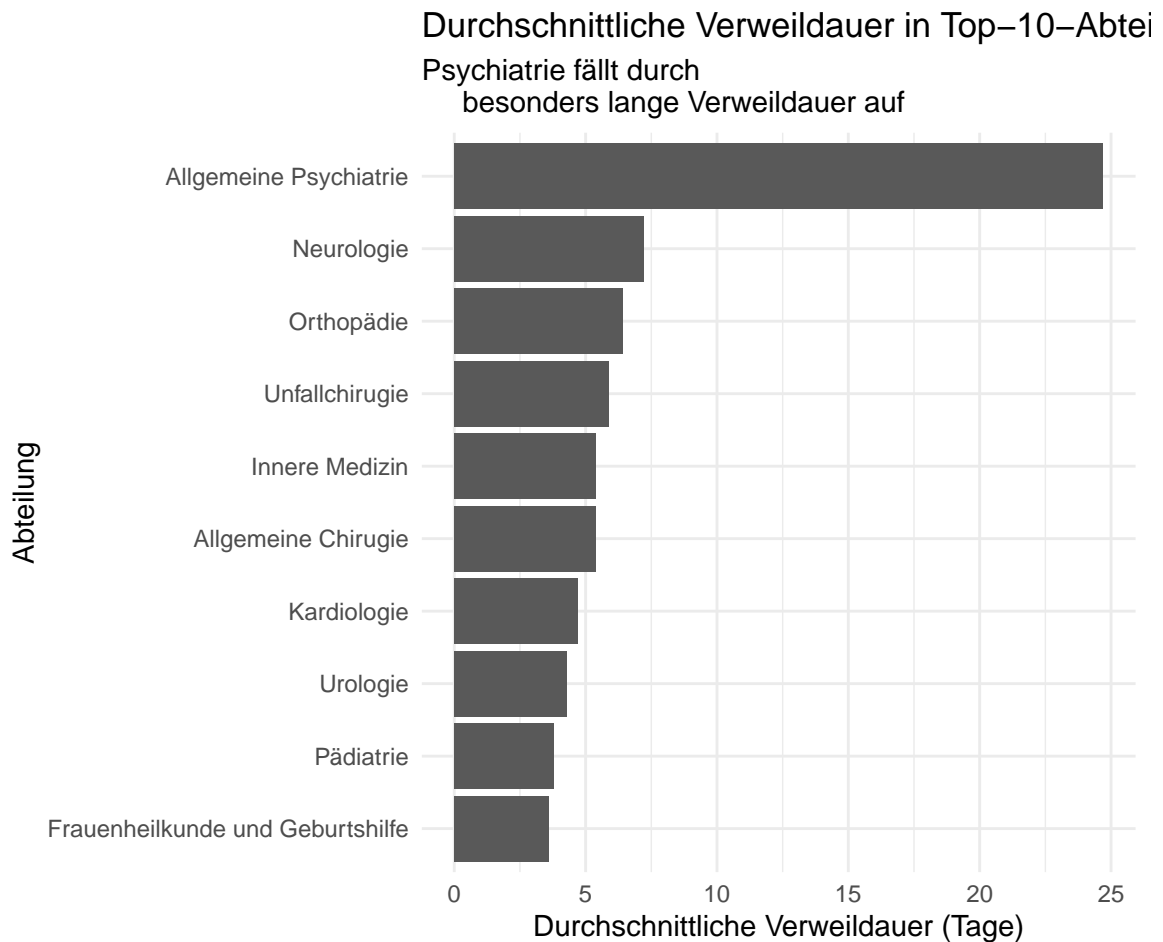
下图显示了排名前 10 位的科室的平均住院日数，并突出显示了精神科病房的特殊价值：

```
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()

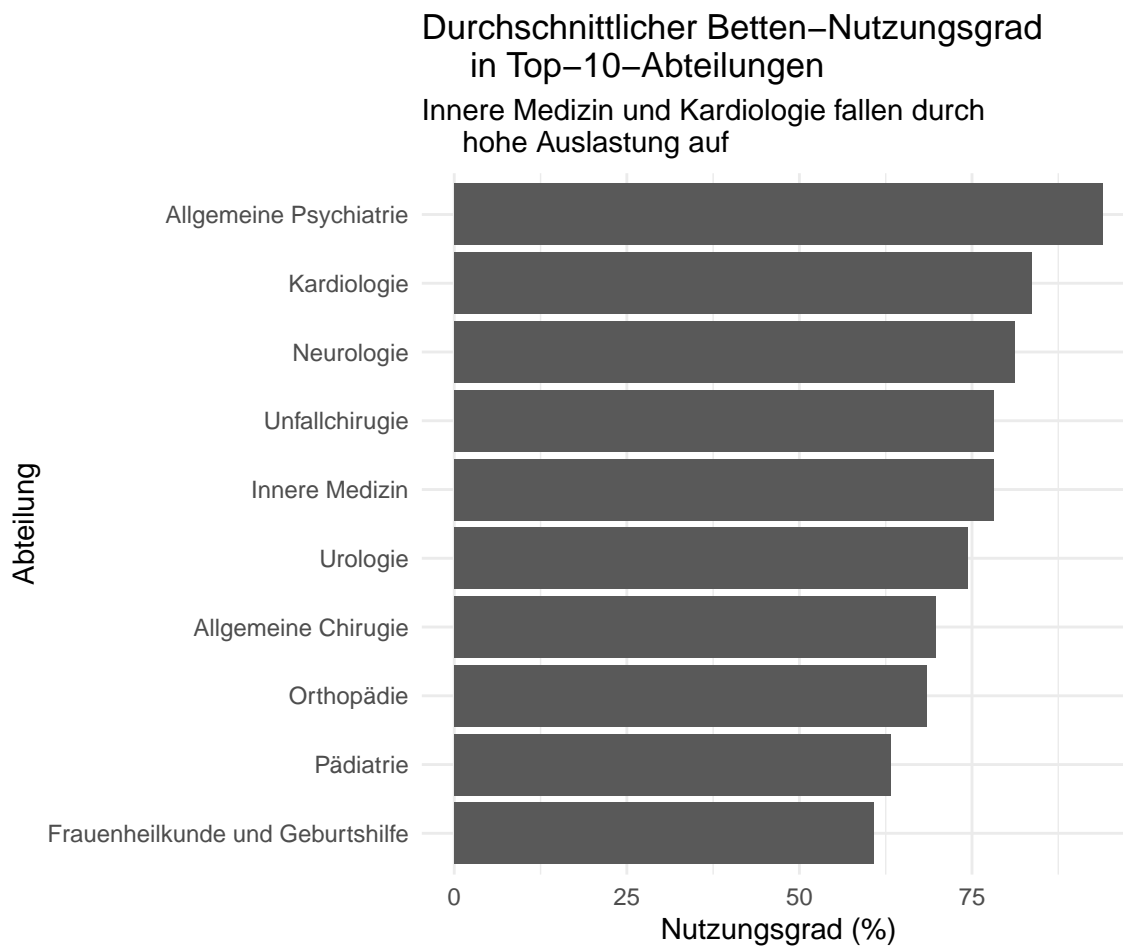
```



为什么内科的病例数如此之高？

综合考虑所有可用变量，我们无法找到导致病例数与床位比率较高的明确原因。有趣的是，心脏病科也表现出类似但不那么明显的关系，这可能是由于这两个学科之间存在重叠。众所周知，内科和心脏病科的护士工作压力相对较大。现有数据无法从统计学上证实这一假设，但它确实提供了一个合理的解释。下图显示了排名前十的科室的平均利用率：

```
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()
```



#### 解读

内科的高利用率可能表明该领域对床位或人员的需求较大。然而，这些数据无法得出具体的行动建议，但可以在未来的规划中讨论。