

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

Patrick Witczak

2025-09-02

Table of contents

Wprowadzenie [Cel]	1
Wybór i import zbioru danych	2
Przegląd zbioru danych	4
Opis zmiennej	4
Rozkład liczby łóżek według oddziałów	4
Rozkład liczby przypadków według oddziałów	5
Analiza poprzez wizualizacje	7
Liczba łóżek i przypadków: Analiza korelacji	7
10 największych oddziałów: 80% liczby przypadków i odpowiadająca jej liczba łóżek	8
Ocena wyników	13
Dlaczego liczba przypadków na oddziałach psychiatrycznych jest tak niska?	13
Dlaczego liczba pacjentów w oddziałach interny jest tak duża?	14
Interpretacja	16

Wprowadzenie [Cel]

Niniejsza analiza bada, jak możliwości opieki w niemieckich szpitalach w 2019 roku odnoszą się do rzeczywistego obciążenia pracą. W tym celu porównano liczbę pacjentów i liczbę łóżek na wszystkich oddziałach szpitalnych. Rok 2019 reprezentuje typowy okres opieki szpitalnej

sprzed pandemii COVID-19. To porównanie ujawnia różnice między dostępnymi możliwościami a zapotrzebowaniem pacjentów. Ponieważ na te różnice mogą wpływać różne czynniki, w celu dalszego pogłębienia analizy wykorzystano dodatkowe zmienne.

Wybór i import zbioru danych

Pierwotnie do analizy przeznaczono oficjalne dane Federalnego Urzędu Statystycznego (Destatis). Ponieważ jednak bezpośredni dostęp jest płatny, wykorzystano ogólnodostępne dane z raportów szpitalnych (Springer Verlag, 2022). Odnoszą się one do oficjalnych danych Destatis i reprezentują wartości z 2019 roku. Ponieważ zbiór danych jest stosunkowo niewielki i obejmuje około 40 oddziałów, import przeprowadzono ręcznie za pomocą funkcji `tribble()` z pakietu `tidyverse`. Umożliwiło to utworzenie zbioru danych i jego dalsze przetwarzanie bezpośrednio w R:

```

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)

```

Przegląd zbioru danych

Opis zmiennej

Zbiór danych zawiera następujące zmienne:

Zmienne główne do analizy

Dept_Beds: Łączna liczba łóżek na oddział

Case Count: Łączna liczba przypadków na oddział

Dodatkowe zmienne do pogłębionej analizy

Dept_Number: Liczba placówek z danym oddziałem

Utilization Rate: Średni wskaźnik obłożenia łóżek (w %)

Avg_Length_of_Stay: Średnia długość pobytu na przypadek (w dniach)/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

Rozkład liczby łóżek według oddziałów

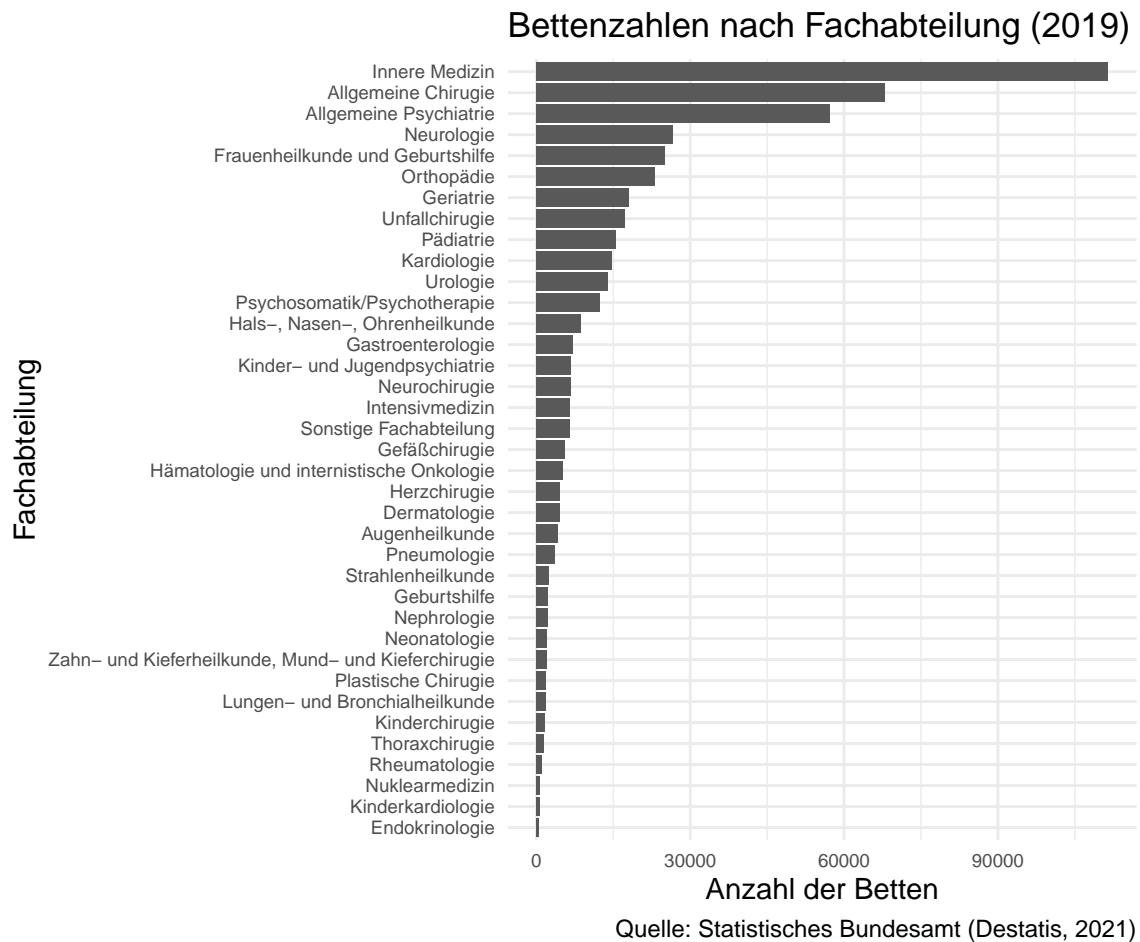
Porównanie liczby łóżek ujawnia silną koncentrację: mniej niż połowa z 37 oddziałów odpowiada za około 80% wszystkich łóżek szpitalnych. Podkreśla to kluczowe znaczenie głównych specjalizacji podstawowych, takich jak interna czy chirurgia ogólna, podczas gdy specjalizacje wysokospecjalistyczne mają naturalnie znacznie mniejszą liczbę łóżek.

```
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))

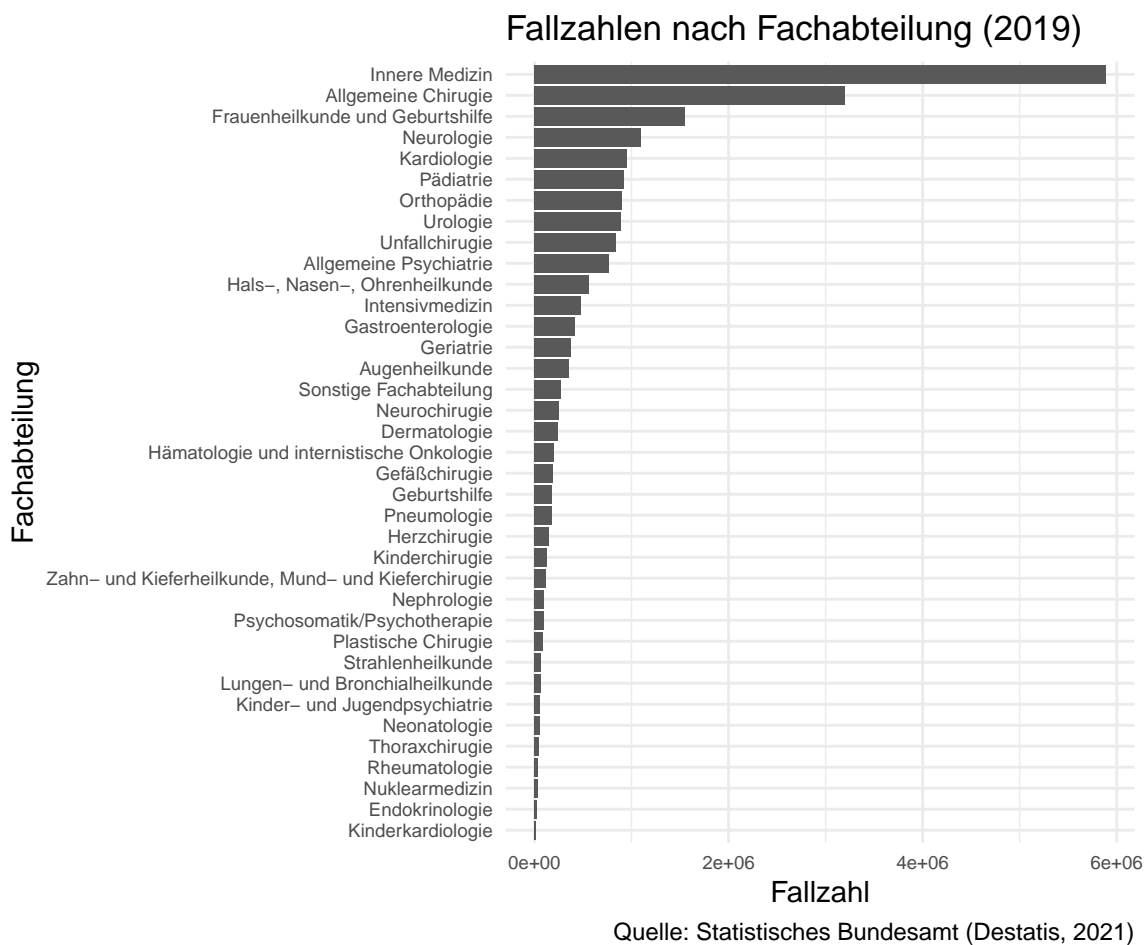
```



Rozkład liczby przypadków według oddziałów

Rozkład liczby przypadków wykazuje podobny wzorec jak liczba łóżek: znaczna część przypadków koncentruje się na kilku dużych oddziałach. Na pierwszy rzut oka te dwie zmienne wydają się być ściśle powiązane, chociaż występują odchylenia w poszczególnych oddziałach. Zależność ta zostanie zbadana w kolejnym etapie analizy za pomocą wykresu punktowego.

```
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))
```



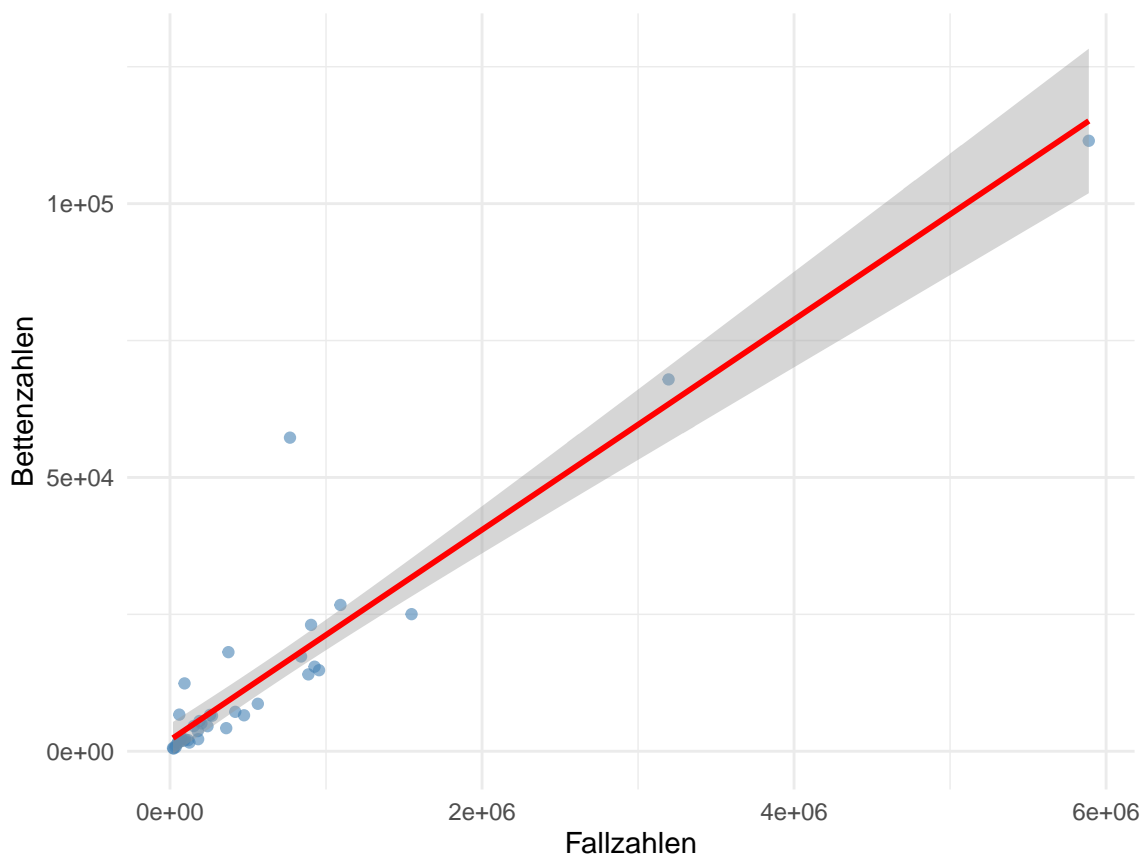
Analiza poprzez wizualizacje

Liczba łóżek i przypadków: Analiza korelacji

Analiza wskazuje na dodatnią zależność liniową między liczbą przypadków a liczbą łóżek: Oddziały z większą liczbą przypadków mają zazwyczaj więcej łóżek. Pomimo tej ogólnej liniowości, poszczególne oddziały wykazują odchylenia od linii trendu. Ze względu na niewielką liczbę 37 obserwacji, interpretacje powinny być ostrożne. Dalsze analizy mogą pomóc w zbadaniu możliwych przyczyn tych odchyleń.

```
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()
```

Beziehung zwischen Fallzahlen und Bettenzahlen (2019)



Quelle: Statistisches Bundesamt (Destatis, 2021)

10 największych oddziałów: 80% liczby przypadków i odpowiadająca jej liczba łóżek

Problemy

Celem analizy było bezpośrednie porównanie liczby przypadków i liczby łóżek na niemieckich oddziałach szpitalnych. Pojawiły się następujące wyzwania:

Dane zawierają wiele wysoce wyspecjalizowanych oddziałów z niską liczbą przypadków, co utrudnia wizualizację.

Pierwotna struktura zbioru danych nie nadaje się do bezpośredniego porównania dwóch kluczowych danych.

Liczba przypadków i liczba łóżek różnią się znacznie pod względem wielkości, co wymaga normalizacji.

Te problemy doprowadziły do zastosowania **logiki Pareto** i **restrukturyzacji zbioru danych**.

Logika Pareto

Aby zmniejszyć liczbę wyświetlanych oddziałów, zachowując jednocześnie reprezentatywne wyniki, do obciążenia pracą zastosowano logikę Pareto. Wybrano tylko największe oddziały, które łącznie stanowią 80% obciążenia pracą. Pozwoliło to zmniejszyć liczbę oddziałów z 37 do 10, zachowując przejrzystość wizualizacji i podkreślając najważniejsze działy.

```
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 <chr>	Fallzahl <dbl>	rel_cumsum_Fallzahl <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 Zmiana struktury zbioru danych

Z punktu widzenia wizualizacji, umieszczanie liczby łóżek i liczby przypadków w osobnych kolumnach jest niepraktyczne. Dlatego te dwie zmienne numeryczne zostały przekształcone w zmienną numeryczną (wartość) i zmienną kategorialną (wartość kluczową):

Wartość kluczowa zawiera nazwę oryginalnej zmiennej („Dept_Beds” lub „Case count”).

Wartość zawiera odpowiadające jej liczby.

```
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
```

Ta transformacja umożliwia wyświetlenie obu kluczowych danych na jednym wykresie.

03 Relatywizacja liczb

Ponieważ liczba łóżek odzwierciedla wartości pięciocyfrowe, a liczba przypadków – siedmiocyfrowe, bezpośrednie porównanie nie ma sensu. Dla zapewnienia porównywalności wartości zostały przeskalowane do 100% za pomocą funkcji `geom_col(position = "fill")`, tak aby słupki przedstawiały względne proporcje poszczególnych działów.

Analiza

Stosunki między liczbą przypadków a liczbą łóżek można bezpośrednio porównać w postaci **dwóch nałożonych na siebie wykresów słupkowych 100%**. Taka reprezentacja wyjaśnia, które oddziały są nieproporcjonalnie lub niedostatecznie wyposażone w łóżka i dostarcza wstępnych wskazówek dotyczących potencjalnych problemów w planowaniu przepustowości.

Chociaż większość oddziałów wykazuje niemal identyczne względne udziały w obu wskaźnikach, następujące obserwacje odstające są szczególnie uderzające:

Psychiatria ogólna, gdzie względna liczba łóżek jest około trzy razy większa niż względna liczba przypadków.

Interna, gdzie względna liczba przypadków jest około 5% wyższa niż liczba łóżek.

```

### 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) +

```

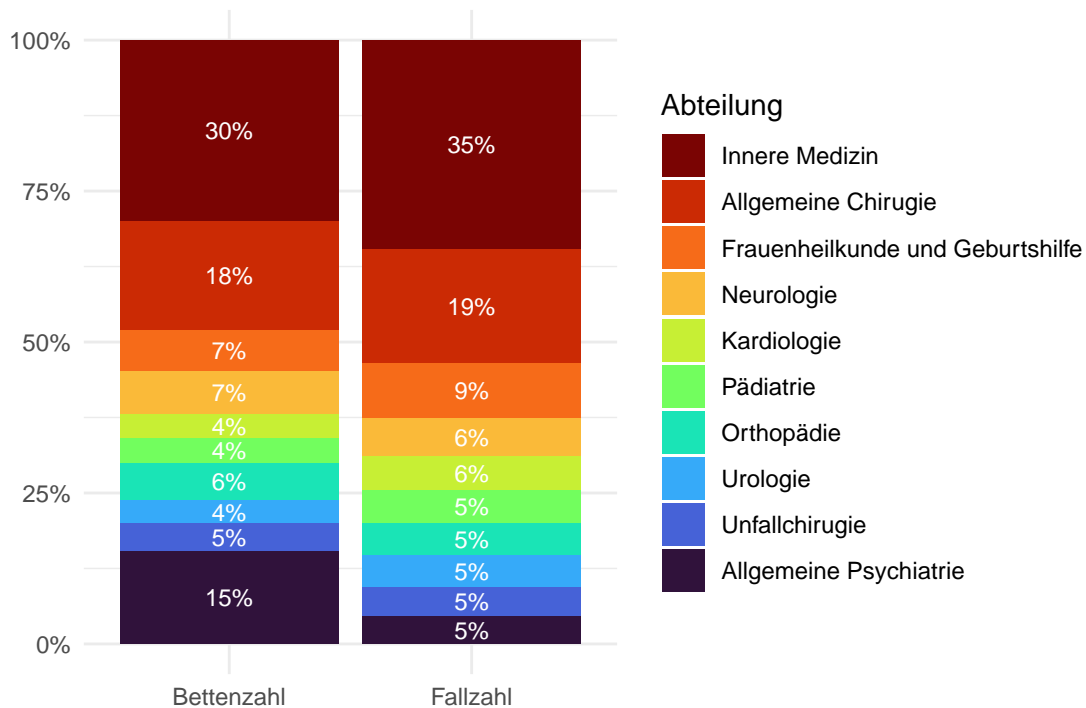
```

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)

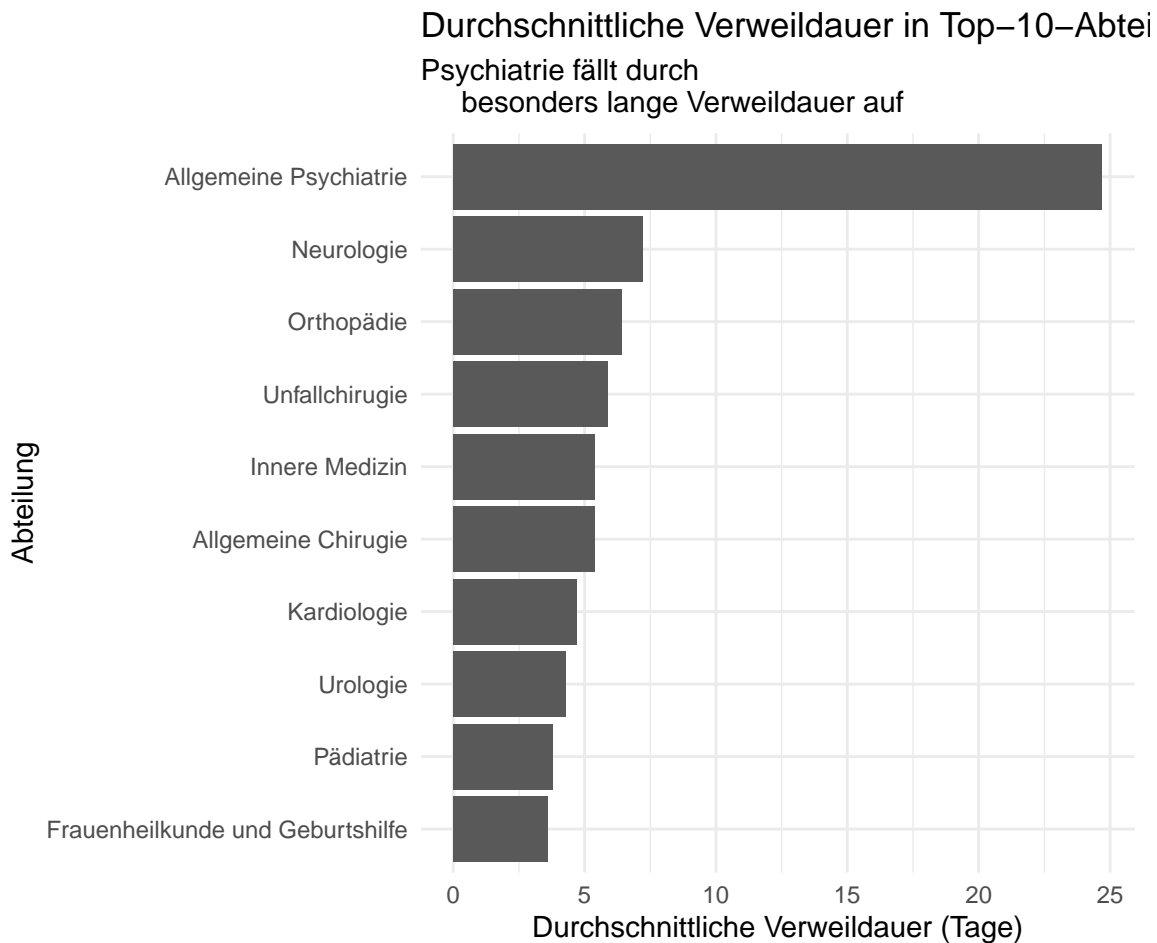
Ocena wyników

Dlaczego liczba przypadków na oddziałach psychiatrycznych jest tak niska?

Przyjrzenie się średniemu czasowi pobytu pozwala na wyjaśnienie: Na oddziałach psychiatrycznych pacjenci przebywają średnio prawie 25 dni, podczas gdy na większości innych oddziałów średnia wynosi około 5 dni. Dłuższe obłożenie prowadzi do mniejszej rotacji pacjentów, co wyjaśnia stosunkowo niską liczbę przypadków.

Poniższy wykres przedstawia średni czas pobytu na 10 największych oddziałach i podkreśla szczególne walory oddziałów psychiatrycznych:

```
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()
```

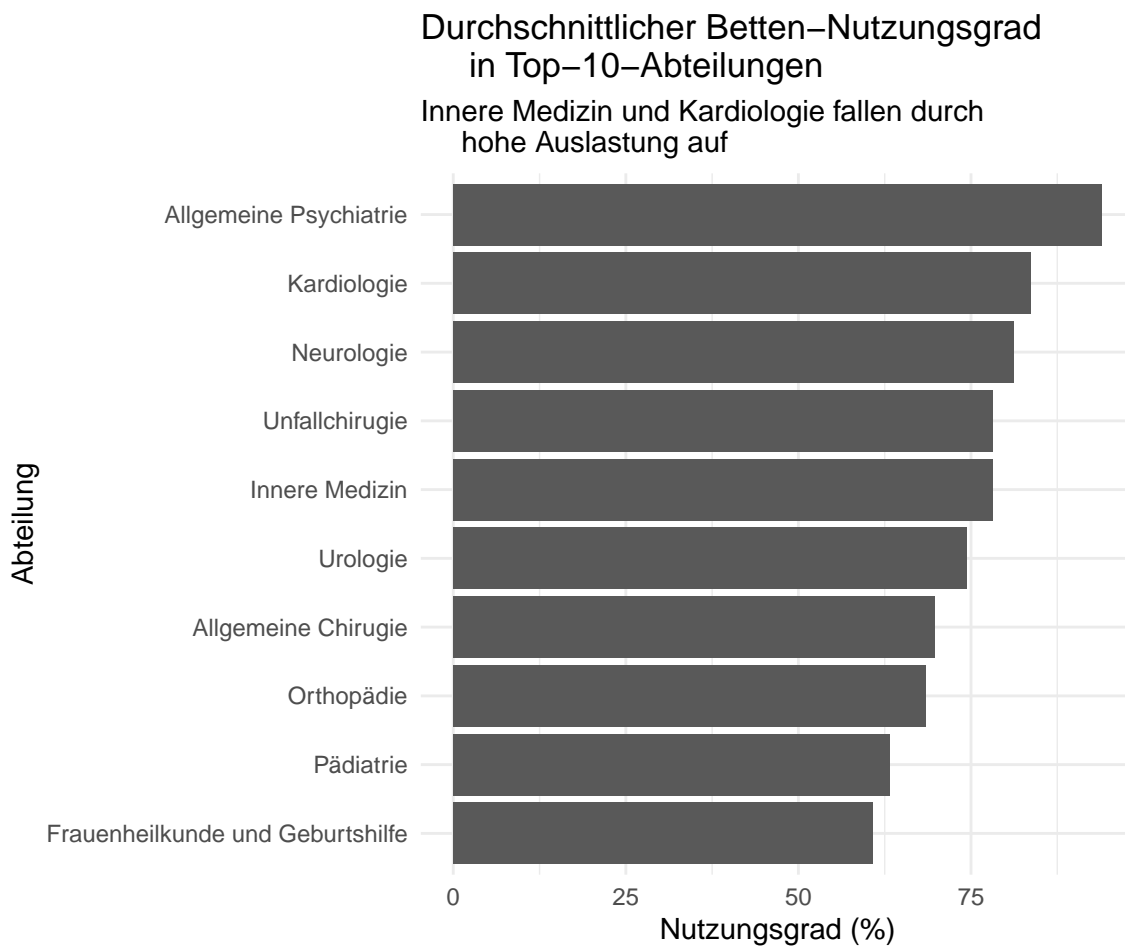


Dlaczego liczba pacjentów w oddziałach interny jest tak duża?

Biorąc pod uwagę wszystkie dostępne zmienne, nie można wskazać jednoznacznych przyczyn wyższego wskaźnika liczby pacjentów do liczby łóżek. Co ciekawe, kardiologia wykazuje podobną, choć mniej wyraźną, zależność, prawdopodobnie ze względu na wzajemne przenikanie się obu dyscyplin. Wiadomo, że pielęgniarki w oddziałach interny i kardiologii pracują pod stosunkowo dużą presją. Tego założenia nie da się uzasadnić statystycznie dostępnymi danymi, ale daje ono wiarygodny kontekst. Poniższy wykres przedstawia średni wskaźnik wykorzystania 10 największych oddziałów:

```
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()
```



Interpretacja

Wysoki wskaźnik wykorzystania w oddziałach chorób wewnętrznych może wskazywać na większe zapotrzebowanie na łóżka lub personel w tym obszarze. Jednak na podstawie tych danych nie można sformułować konkretnych zaleceń dotyczących działań, ale można je omówić w toku dalszego planowania.