Emergency Department Utilization Patterns: A Data Analysis Case Study

Mutsumi Cornett
November 23, 2025

1. Executive Summary

This case study analyzes Emergency Department (ED) utilization patterns in the United States using publicly available national datasets containing approximately 6,700 aggregated records. The goal of this project is to identify which diagnoses, age groups, and time trends drive ED demand and to translate these findings into operational insights for hospitals.

Using RStudio for data cleaning, transformation, and visualization, this analysis answers four key questions:

1. Which diagnoses account for the highest number of ED visits?
2. Which age groups utilize the ED most frequently?
3. How have ED visits changed over time?
4. What insights can hospitals use to improve staffing, resource allocation, and patient flow?

The findings show that symptom-based presentations and injury-related encounters dominate ED volume, that utilization varies across age groups, and that yearly ED visit totals remain consistently high. These results demonstrate how data analytics can support efficient ED operations and help hospitals anticipate future demand.

Based on these trends, EDs may benefit from maintaining adequate staffing levels across triage, acute care, and diagnostic support roles, ensuring rapid access to laboratory and imaging services, and planning for sufficient bed capacity to accommodate consistently high visit volumes.

---

2. Background & Purpose

Emergency Department (ED) utilization is a critical metric for hospital operations and population health. In the years following the COVID-19 pandemic, EDs have experienced capacity strain, staffing shortages, and fluctuating patient acuity. Understanding who uses the ED and why is essential for strategic planning and resource management. It is important to note that the data used in this case study consists of **aggregate national estimates**, not patient-level records.

Hospitals analyze ED utilization data to:

- Forecast staffing needs
- Allocate bed and resource capacity
- Plan for high-demand patient populations
- Improve workflow efficiency
- Support patient and provider safety

---

3. Dataset Description

This analysis uses publicly available national emergency department (ED) utilization data from the ***Centers for Disease Control and Prevention (CDC)****,* ***National Center for Health Statistics (NCHS)****. The dataset originates from the CDC's Emergency Department Visits in the United States database and includes annual estimates of ED visits across major clinical categories.*

Dataset Name

Emergency Department Visits in the United States, 2016–2022

Source

Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS).
Retrieved from the CDC ED Visits Data Portal:
https://www.cdc.gov/nchs/dhcs/ed-visits/index.htm

Timeframe Covered

2016 through 2022 (7 years)

Number of Records (Rows)

The dataset contains approximately ***6,700 rows****, representing combinations of:*

- *Year*
- *Measure type (e.g., diagnosis category, age group, reason for visit)*
- *Subgroup*
- *Estimate*
- *Confidence intervals and standard error values*

Key Variables Used in This Analysis

- ***year** — Reporting year*
- ***measure** — Diagnosis category or reason for visit*
- ***measure_type** — “By age,” “By diagnosis,” “By reason,” etc.*
- ***group** — Broad classification category*
- ***subgroup** — Age groups or diagnostic subcategories*
- ***estimate** — National estimated number of ED visits*
- ***standard_error** — SE of the national estimate*
- ***lower_95_percent_ci / upper_95_percent_ci** — Confidence interval bounds*
- ***reliable** — Indicates whether the estimate meets reliability standards*

Data Quality Considerations

- ***No missing values** were present in the dataset; all variables were fully observed.*
- *Some fields contained **zero values**, but in this dataset a zero does **not** represent missing data. Instead, it is used for categories such as:*
    - *“All diagnoses” (combined totals)*
    - *Categories not ranked within a top 10 list*
- *The dataset provides **annual totals only—**there is no daily, weekly, or hourly visit-level detail.*
- *Individual patient-level data are **not included**; this is an aggregate national dataset designed for public health surveillance.*

---

4. Methodology

Tools Used

- RStudio
- dplyr / tidyverse
- ggplot2
- scales
- janitor

Data Cleaning Process

- Verified dataset contained no missing values using `colSums(is.na(df))`
- Interpreted zero values in *leading_10_ranking*
  → Confirmed that 0 = a meaningful category (e.g., all diagnoses), not NA
  → Therefore, zeros were intentionally retained
- Filtered records by age group and diagnosis category
- Aggregated visit estimates using `summarise()`
- Converted raw counts into millions/billions for visualization clarity
- Exported high-resolution plots for inclusion in report

This ensures the analysis is accurate and aligned with the dataset's coding structure and the workflow adheres to reproducible research principles.

---

5. Analysis

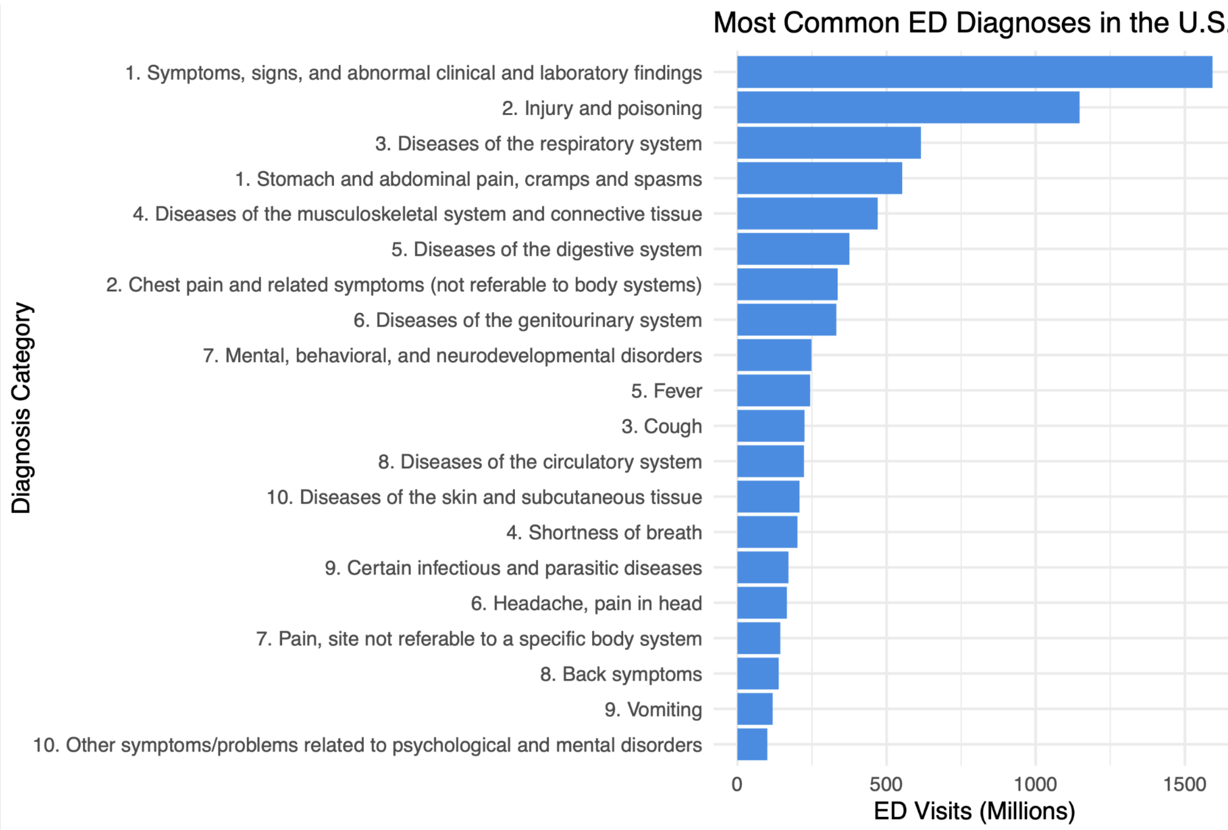
A. Top Emergency Department Diagnoses

Using data from 2016–2022, I aggregated ED visit estimates across all diagnosis categories. After excluding "All diagnoses" and "All reasons," the analysis reveals:

- Symptoms, signs, and abnormal clinical/lab findings produced the highest ED volume.
- Injury and poisoning ranked second.
- Respiratory diseases, abdominal pain, and musculoskeletal disorders also contributed substantially.

Symptom-based categories are broad by design and capture diverse patient presentations. This explains why they form the largest category.

These findings show that EDs must maintain strong triage capability for broad, non-specific symptom presentations and acute injuries.



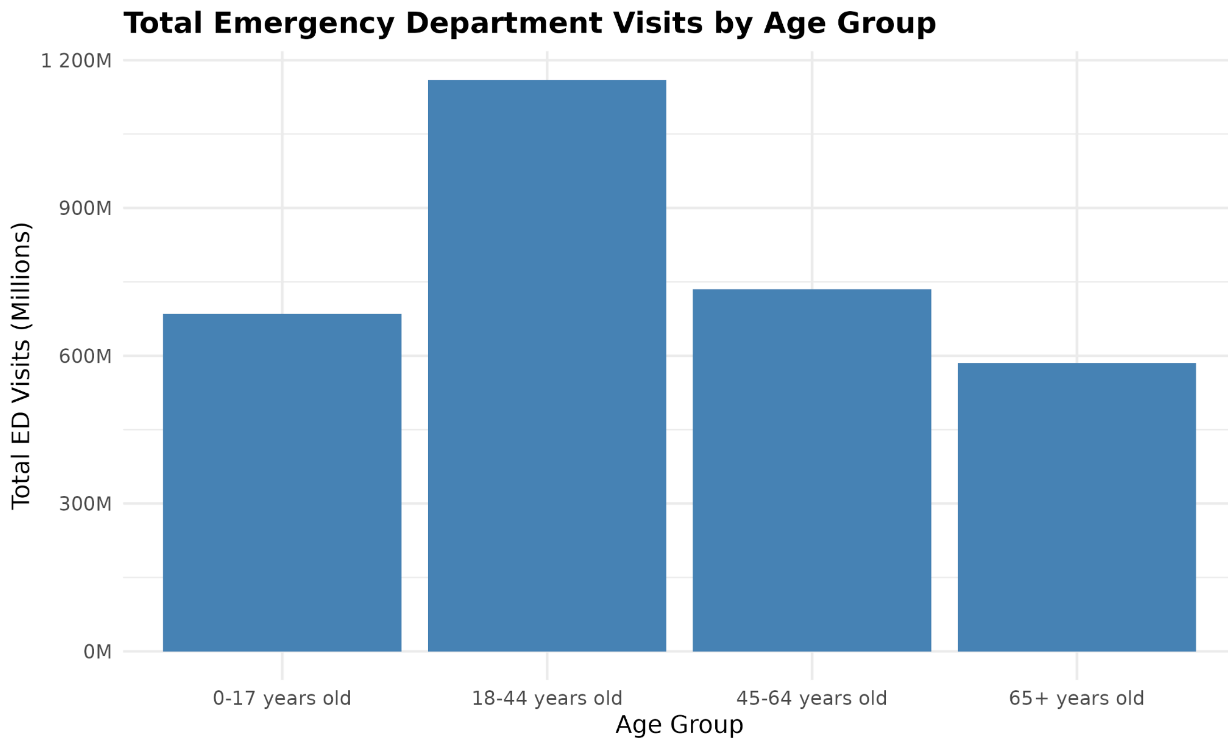
B. ED Visits by Age Group

The dataset includes four standard age groups commonly used in national ED reporting: 0–17, 18–44, 45–64, and 65+. After filtering the dataset to include only age-group-specific records, I summarized total ED visits among the four age categories. Age group findings help EDs identify which populations may require specialized staffing (e.g., pediatric emergency nurses, geriatric specialists).

A bar chart comparing total visits shows:

- Which age group has the highest utilization
- How ED demand varies across the lifespan
- Which populations drive emergency care demand

This supports targeted resource and staffing allocation.



C. ED Visit Trends Over Time

ED visits were aggregated by year (2016–2022). A line chart shows:

- National ED visits remain consistently high
- Total yearly volume ranges from 3.1 to 3.3 billion visits across the 7-year period
- Minor fluctuations occur from year to year
- No evidence of decline in overall utilization
- COVID-19 likely influenced some fluctuations, though this dataset does not break out COVID-specific visits.

This trend underscores the necessity for stable ED staffing and infrastructure.

6. Insights for Hospitals

Key operational takeaways:

- High symptom-based visit volume suggests need for strong triage and diagnostic capacity
- Injury-related visits highlight the importance of trauma resources
- The most frequent age groups can help target staffing for pediatric, adult, or geriatric care
- Stable yearly visits indicate ED demand is not decreasing despite outpatient care expansion
- Predictable patterns enable more efficient scheduling and budgeting

7. Limitations

- Dataset lacks time-of-day or day-of-week detail
- Cannot analyze peak hours or seasonal trends
- Diagnoses are grouped categories, not ICD-10 code-level detail
- Certain real-world factors (acuity, admit rate, length of stay) are not included
- The dataset provides national-level estimates only; results cannot be used to infer hospital-specific performance.

---

8. Conclusion

This project demonstrates how national ED utilization data can be transformed into actionable operational insights. By identifying high-volume diagnoses, key demographic groups, and yearly trends, hospitals can better anticipate resource needs and optimize emergency care delivery. This case study also demonstrates how publicly available data combined with analytical tools such as R can generate meaningful insights that support evidence-based decision-making in healthcare operations. The analysis also highlights the value of reproducible code-based workflows for healthcare analytics.

---

9. Appendix A

A1. Load Packages and Import the Dataset

```
# Load required packages
library(tidyverse)
library(janitor)
library(scales)

# Import dataset
df <- read_csv("cdc_ed_visits_2016_2022.csv") %>%
  clean_names()
```

A2. Inspect Data Structure

```
# View column names and structure
names(df)
str(df)

# Check for missing values
colSums(is.na(df))
```

A3. Data Cleaning Notes

- No missing values
- Zero values in ranking variables are meaningful, not NA
  No recoding was required.

Section B — Diagnosis Category Analysis

B1. Filter to Specific Diagnosis Categories

```
# Remove “All diagnoses” and “All reasons” so results are meaningful.

df_specific <- df %>%
  filter(!measure %in% c("All diagnoses", "All reasons"))
```

B2. Aggregate Diagnosis-Level ED Visit Counts

```
df_diagnosis <- df_specific %>%
  group_by(measure) %>%
  summarise(total_visits = sum(estimate, na.rm = TRUE)) %>%
  ungroup()
```

B3. Convert Visit Counts into Millions

```
df_diagnosis <- df_diagnosis %>%
  mutate(total_visits_millions = total_visits / 1e6)
```

B4. Diagnosis Bar Chart

```
ggplot(df_diagnosis,
       aes(x = reorder(measure, total_visits_millions),
           y = total_visits_millions)) +
   geom_col(fill = "#4A90E2") +
   coord_flip() +
   labs(
      title = "Most Common Emergency Department Diagnoses (2016–2022)",
      x = "Diagnosis Category",
      y = "ED Visits (Millions)"
   ) +
   theme_minimal() +
   theme(
      axis.text.y = element_text(size = 9),
      axis.text.x = element_text(size = 9)
   )
```

B5. Save Diagnosis Chart

```
ggsave(
   filename = "ed_diagnoses_bar_chart.png",
   plot = last_plot(),
   width = 10,
   height = 7,
   dpi = 300
)
```

Section C — Age Group Analysis

C1. Filter to Age-Specific Rows

```
df_age <- df %>%
  filter(group == "By age")
```

C2. Summarize ED Visits by Age Group

```
age_summary <- df_age %>%
  group_by(subgroup) %>%
  summarise(total_visits = sum(estimate, na.rm = TRUE))
```

C3. Age Group Bar Chart

```
ggplot(age_summary, aes(x = subgroup, y = total_visits)) +
  geom_col(fill = "steelblue") +
  scale_y_continuous(labels = label_number(scale = 1e-6, suffix = "M")) +
  labs(
    title = "Total Emergency Department Visits by Age Group",
    x = "Age Group",
    y = "Total ED Visits (Millions)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 0, hjust = 0.5),
    plot.title = element_text(face = "bold")
  )
```

C4. Save Age Group Chart

```
ggsave("ED_visits_by_age_group.png", width = 10, height = 6, dpi = 300)
```

Section D — Yearly Trend Analysis (2016–2022)

D1. Aggregate Total ED Visits by Year

```
df_year <- df %>%
  group_by(year) %>%
  summarise(total_visits = sum(estimate, na.rm = TRUE))
```

D2. Convert to Billions for Readability

```
df_year <- df_year %>%
  mutate(total_visits_billion = total_visits / 1e9)
```

D3. Yearly Line Chart

```
ggplot(df_year, aes(x = year, y = total_visits_billion)) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  labs(
    title = "Yearly Emergency Department Visits (2016–2022)",
    x = "Year",
```

```
    y = "ED Visits (Billions)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(face = "bold")
  )
```

D4. Save Yearly Line Chart

```
ggsave("ED_visits_over_time.png", width = 10, height = 6, dpi = 300)
```