PHW251 Take-home Midterm Exam

AUTHOR
Val Stacey

PUBLISHED
October 26, 2025

```
library(readr)
library(here)
library(janitor)
library(dplyr)
library(tidyr)
library(knitr)
library(kableExtra)
```

NOTE: all files to support this take-home exam can be found in my GitHub repo: https://github.com/EvalstacE/PHW251_R/tree/main/midterm

Question 1

```
data <- read.csv(file = "data/exam_data_2025_version1.csv")
```

1A:

- Year: **integer**
- OSHPD ID: **integer**
- Facility Name: **character**
- County Name: **character**
- ER Service Level Desc: **character**
- Type: **character**
- Count: **integer**

supporting code:

```
str(data)
```

Question 2

```
data <- data %>% clean_names() %>%
  rename("count_encounters" = "count")

# Note: I renamed 'count' to 'count_encounters' to be safe --
## -- since 'count' is typically reserved and could cause issues later
## -- (same name as base R function)
```

2A:

- year
- oshpd_id
- facility_name
- county_name
- er_service_level_desc
- type
- count_encounters

supporting code:

```
colnames(data)
```

Question 3

```
q3_df <- data %>%
  filter(er_service_level_desc == "BASIC",
         year %in% 2014:2018)
```

3A: — 2,050 records

supporting code:

```
nrow(q3_df)
```

Question 4

```
q3_df <- q3_df %>%
  group_by(oshpd_id, year) %>%
  mutate(total_encounters = sum(count_encounters)) %>%
  ungroup()
```

4A: — 115,702 total encounters

supporting code:

```
q3_df %>%
  filter(
    oshpd_id == "106304409",
    year == "2017"
  ) %>%
  pull(total_encounters)%>%
  first()
```

4B: — 19.3%

supporting code:

```
q3_df <- q3_df %>%
  mutate(
    pct_encounter_type =
      round(100 * count_encounters / total_encounters, 1)
    )
```

```
q3_df %>%
  filter(
    oshpd_id == "106301175",
    year == "2015",
    type == "ED_Admit"
  ) %>%
  pull(pct_encounter_type)
```

Question 5

```
q5_df <- q3_df %>%
  filter(
    type == "ED_Admit"
  )
```

5A:

```
q5_df <- q5_df %>%
  arrange(year, desc(pct_encounter_type))
```

Subset to demonstrate correct arrangement:

Year	Facility	Type	% Encounters
2014	COLLEGE MEDICAL CENTER	ED_Admit	30.4
2014	NORWALK COMMUNITY HOSPITAL	ED_Admit	29.8
2015	LOS ANGELES COMMUNITY HOSPITAL AT BELLFLOWER	ED_Admit	100.0
2015	COLLEGE MEDICAL CENTER	ED_Admit	34.0
2016	LOS ANGELES COMMUNITY HOSPITAL AT BELLFLOWER	ED_Admit	100.0
2016	GARDENS REGIONAL HOSPITAL AND MEDICAL CENTER	ED_Admit	40.1
2017	GARDENS REGIONAL HOSPITAL AND MEDICAL CENTER	ED_Admit	40.8
2017	COLLEGE MEDICAL CENTER	ED_Admit	38.4
2018	LOS ANGELES COMMUNITY HOSPITAL AT BELLFLOWER	ED_Admit	100.0
2018	MISSION COMMUNITY HOSPITAL - PANORAMA CAMPUS	ED_Admit	40.7

5B:

```
q5_df %>% select(facility_name) %>%
  slice_head(n = 5)
```

Table: Top 5 Facilities by ED Admission Percentage (2014)

Facilities:
COLLEGE MEDICAL CENTER
NORWALK COMMUNITY HOSPITAL
GLENDORA COMMUNITY HOSPITAL
ADVENTIST HEALTH GLENDALE
EMANATE HEALTH INTER-COMMUNITY HOSPITAL

supporting code for table:

```
q5_top <- q5_df %>%
  select(facility_name) %>%
  slice_head(n = 5)

q5_top_tbl <- q5_top %>%

kable(format = "html", align = "l", col.names = c("Facilities:")) %>%
kable_styling(
    bootstrap_options = c("striped", "condensed", "bordered"),
    position = "center") %>%
row_spec(0,
  bold = TRUE,
  color = "#ffffff",
  background = "#404682",
  extra_css = "padding: 8px 12px;font-size: 24px!important;") %>%
row_spec(seq(1,nrow(q5_top),2),
  background="#e3e3ec") %>%
column_spec(1,
  extra_css = "padding: 8px 44px;font-size: 20px!important;")
```

Question 6

```
avg_pct = mean(q5_df$pct_encounter_type, na.rm = TRUE)

q5_df <- q5_df %>%
  mutate(
    above_below_avg =
      case_when(
        pct_encounter_type >= avg_pct ~ "above",
        TRUE ~ "below"
        )
      )
```

6A: — 14.15%

6B: — below

supporting code:

```
q5_df %>%
  filter(
    oshpd_id == "106190949",
    year == "2017"
  ) %>%
  pull(above_below_avg)
```

Question 7

```
q7_df <- q5_df %>%
  filter(county_name %in% c("SAN LUIS OBISPO", "SANTA CLARA", "ORANGE"))
```

7A: — 172 records

supporting code:

```
nrow(q7_df)
```

7B: — FOOTHILL REGIONAL MEDICAL CENTER

(33.9% in 2017)

note: this facility also has the highest overall average across all years at 31.3%

supporting code:

```
q7_df <- q7_df %>%
  arrange(desc(pct_encounter_type))
```

```
q7_df %>% select(facility_name, year, pct_encounter_type) %>%
  slice_head(n = 1)

## I was not sure if you were looking for highest overall,
### -- or the facility with the average highest.
#### --- (turns out they are the same anyway...)

## To find the facility with the highest average, you could run:
q7b <- q7_df %>% select(year, facility_name, pct_encounter_type) %>%
  group_by(facility_name) %>%
  summarise(avg_facility_pct = mean(pct_encounter_type, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(avg_facility_pct))

q7b %>% slice_head(n =1)
```

Question 8

```
q8_df <- data %>%
  filter(type == "ED_Visit", year == "2018") %>%
  select(facility_name, county_name,
         er_service_level_desc, type,
         count_encounters)
```

8A: — HOAG MEMORIAL HOSPITAL PRESBYTERIAN

(98,209 total encounters in 2018)

supporting code:

```
q8_df <- q8_df %>%
  arrange(desc(count_encounters))

q8_df %>% slice(10)
```

Question 9

9A: — 128 counties in “low” utilization category

Table: Number of Counties by ER Utilization Category (2018)

ER Utilization Category	Number of Counties
High	**0**
Medium	**38**
Low	**128**
Very Low	**76**

supporting code (creating categories):

```
q9_sum_df <- q8_df %>%

mutate(
    visit_category =
       case_when(
         count_encounters > 178649 ~ "High",
         between(count_encounters, 66522, 178649) ~ "Medium",
         between(count_encounters, 22027, 66521) ~ "Low",
         count_encounters <=  22026 ~ "Very Low",
         TRUE ~ NA_character_
      ),

   visit_category =
     factor(visit_category,
            levels = c("High", "Medium", "Low", "Very Low")
            )
  ) %>%

group_by(visit_category, .drop = FALSE) %>%
  summarise(total_cntys = n()) %>%
  ungroup()
```

supporting code for table:

```
q9_sum_tbl <- q9_sum_df  %>%

kable(
    format = "html",
    align = "cc",
    col.names = c("ER Utilization Category", "Number of Counties")
) %>%

kable_styling(
    bootstrap_options = c("striped", "condensed"),
    position = "center"
)%>%
column_spec(1, extra_css = "padding: 8px 12px;font-size:16px;") %>%
column_spec(2, bold = TRUE, extra_css = "padding: 8px 12px;font-size:16px;") %>%
row_spec(0, bold = TRUE, extra_css = "padding: 8px 12px;font-size:
         18px!important;")
```

Question 10

10A. — 206 records after pivot

10B:

Function used from `tidyr` : — `pivot_wider()`

arguments used:

- `id_cols =`
- `names_from =`
- `values_from =`
- `values_fill =`

supporting code:

```
q10_df <- data %>%

  filter(
      year == "2013",
      er_service_level_desc == "BASIC"
  ) %>%

  select(
    facility_name, county_name,
    type, count_encounters
  ) %>%

  pivot_wider(
    id_cols = c(facility_name, county_name),
    names_from = type,
    values_from = count_encounters,
    values_fill = 0
  )
```

Question 11 (extra credit)

Bring in required libraries:

```
library(readr)
library(here)
library(janitor)
library(dplyr)
library(tidyr)
library(knitr)
library(kableExtra)
library(reactable)
library(htmltools)
```

Bring in original df and clean column names:

```
data <- read.csv(file = "data/exam_data_2025_version1.csv") %>%
  clean_names() %>%
  rename("count_encounters" = "count")
```

Create the summary table:

```
q11_df <- data %>%

filter(type == "ED_Visit", year == "2018") %>%

select(
    facility_name, county_name,
    er_service_level_desc, type,
    count_encounters
) %>%

mutate(
    visit_category =
       case_when(
         count_encounters > 178649 ~ "High",
         between(count_encounters, 66522, 178649) ~ "Medium",
         between(count_encounters, 22027, 66521) ~ "Low",
         count_encounters <=  22026 ~ "Very Low",
         TRUE ~ NA_character_
      ),

   visit_category =
     factor(visit_category,
            levels = c("High", "Medium", "Low", "Very Low")
            )
) %>%

group_by(visit_category, .drop = FALSE) %>%
  summarise(total_cntys = n()) %>%
  ungroup()
```

Question 12 (extra credit)

Table A:

(with `kable` and `kableExtra` packages)

Number of Counties by ER Utilization Category (2018):

ER Utilization Category	Number of Counties
High	**0**
Medium	**38**
Low	**128**
Very Low	**76**

ER Utilization Categories (# of ER Visits):

"High": > 178,649; "Medium": > 66,521; "Low": > 22,026; "Very Low": ≤ 22,026;

Table B:

(with `reactable` and `htmltools` packages)

Number of Counties by ER Utilization Category (2018):

ER Utilization Category	Number of Counties
High	0
Medium	38
Low	128
Very Low	76

ER Utilization Categories (# of ER Visits):

"High": > 178,649; "Medium": > 66,521; "Low": > 22,026; "Very Low": ≤ 22,026;

Code to create Table A (kable):

```
q11_tbl_a <- q11_df %>%

kable(
    format = "html",
    align = "cc",
    col.names = c("ER Utilization Category", "Number of Counties")
) %>%

kable_styling(
    bootstrap_options = c("striped", "condensed", "bordered"),
    position = "center"
) %>%

row_spec(0,
  bold = TRUE,
  color = "#ffffff",
  background = "#404682",
  extra_css = "padding: 8px 12px;font-size: 24px!important;"
) %>%

row_spec(seq(1,nrow(q11_df),2),
  background="#e3e3ec"
) %>%

column_spec(1,
  bold = TRUE,
  extra_css = "padding: 8px 12px;font-size: 20px!important;"
) %>%

column_spec(2,
  bold = TRUE,
  extra_css = "padding: 8px 12px;font-size: 20px!important;"
)
```

Code to create Table B (reactable):

```
max_cnt <- max(q11_df$total_cntys, na.rm = TRUE)

q11_tbl_b <- reactable(
  q11_df,
  columns = list(
    visit_category = colDef(
      name  = "ER Utilization Category",
      align = "center",
      style = list(
        fontWeight = "700",
        fontSize   = "20px",
        padding    = "8px 12px"
      )
    ),
    total_cntys = colDef(
      name  = "Number of Counties",
      align = "center",
      cell = function(value) {
        width <- paste0((value / max_cnt) * 100, "%")
        div(
          style = list(
            display     = "flex",
            alignItems  = "center",
            justifyContent = "flex-start",
            gap         = "10px"
          ),

          tags$span(
            style = "font-weight:700; font-size:20px; text-align:right; min-
          width:60px",
            value
          ),

          div(
            style = list(
              flexGrow        = "1",
              height       = "14px",
              background   = "#f2f2f6",
              border       = "1px solid #404682",
              borderRadius = "4px",
              overflow     = "hidden",
              display      = "inline-block"
            ),
            div(
              style = list(
                width        = width,
```

```
                height       = "100%",
                background   = "#404682",
                borderRadius = "1px"
              )
            )
          )
        )
      },
      style = list(
        fontWeight = "700",
        fontSize   = "20px",
        padding    = "8px 12px"
      )
    )
  ),
  bordered = TRUE,
  striped  = TRUE,
  highlight = FALSE,
  compact  = FALSE,
  theme = reactableTheme(
    headerStyle = list(
      background = "#404682",
      color      = "#ffffff",
      fontWeight = "700",
      fontSize   = "24px",
      padding    = "8px 12px"
    ),

    borderColor = "#d9d9d9",
    borderWidth = "1px",
    cellPadding = "8px 12px",
    stripedColor = "#e3e3ec"
  ),
  wrap = FALSE
)
```