Analysis of gardasil shots by demographic factors

This program reads data on Gardasil vaccinations in young women. Find more information in the data dictionary.

The program was written by Steve Simon on 2024-09-07 and is placed in the public domain.

Load the tidyverse library

For most of your programs, you should load the tidyverse library. The messages and warnings are suppressed.

```
        library(tidyverse)
```

Read the data and view a brief summary

Use the read_csv function to read the data. The glimpse function will produce a brief summary. Use tolower to convert uppercase to lowercase.

```
        gard <- read_csv(
          file="../data/gardasil.csv",
          col_names=TRUE,
          col_types="nnnnnnnnnn")
        names(gard) <- tolower(names(gard))
        glimpse(gard)
```

```
Rows: 1,413
Columns: 10
$ age           <dbl> 21, 21, 20, 14, 17, 11, 17, 15, 13, 18, 17, 22, 16, 13, …
$ agegroup      <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0,…
$ race          <dbl> 0, 0, 0, 0, 3, 1, 0, 3, 3, 0, 1, 0, 3, 1, 1, 0, 1, 1, 0,…
$ shots         <dbl> 3, 3, 1, 3, 2, 1, 1, 3, 3, 3, 2, 2, 1, 2, 1, 1, 1, 3, 3,…
$ completed     <dbl> 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1,…
$ insurancetype <dbl> 3, 3, 1, 3, 3, 0, 3, 1, 1, 2, 1, 3, 1, 3, 0, 1, 1, 1, 1,…
$ medassist     <dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,…
$ location      <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
```

```
$ locationtype  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…
$ practicetype  <dbl> 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1,…
```

Question 7: First create factors for medassist

The factor function identifies a variable as categorical and assigns labels to number codes. You don't necessarily need to use factor if the data you read in is character strings, as R automatically treats those variable as categorical.

```
gard$medassist <- factor(
  gard$medassist,
  levels=0:1,
  labels=c(
    "No medical assistance",
    "Received medical assistance"))
```

Question 7: Summarize and interpret the percentage of patients receiving medical assistance. Be sure to convert the number codes for this variable into labels using the factor function

```
gard |>
  count(medassist) |>
  mutate(total=sum(n)) |>
  mutate(pct=round(100*n/total))
```

```
# A tibble: 2 × 4
  medassist                     n total   pct
  <fct>                     <int> <int> <dbl>
1 No medical assistance      1138  1413    81
2 Received medical assistance  275  1413    19
```

Eighty one percent of patients received at least some medical assistance while the remaining 19% did not.

Create factors for shots

It is a bit silly to replace 1, 2, 3 with One, Two, Three. The main reason is to clearly identify shots as categorical rather than continuous.

```
gard$shots <- factor(
  gard$shots,
  levels=1:3,
  labels=c(
    "One",
    "Two",
    "Three"))
```

Counts and percentages for shots

```
gard |>
  count(shots) |>
  mutate(total=sum(n)) |>
  mutate(pct=round(100*n/total))
```

```
# A tibble: 3 × 4
  shots     n total   pct
  <fct> <int> <int> <dbl>
1 One     440  1413    31
2 Two     436  1413    31
3 Three   537  1413    38
```

Slightly more patients got three shots than one or two shots, but this is still less than half of the patients overall.

Question 8: First calculate the percentages for number of shots received by whether the patient received medical assistance. Interpret this chart.

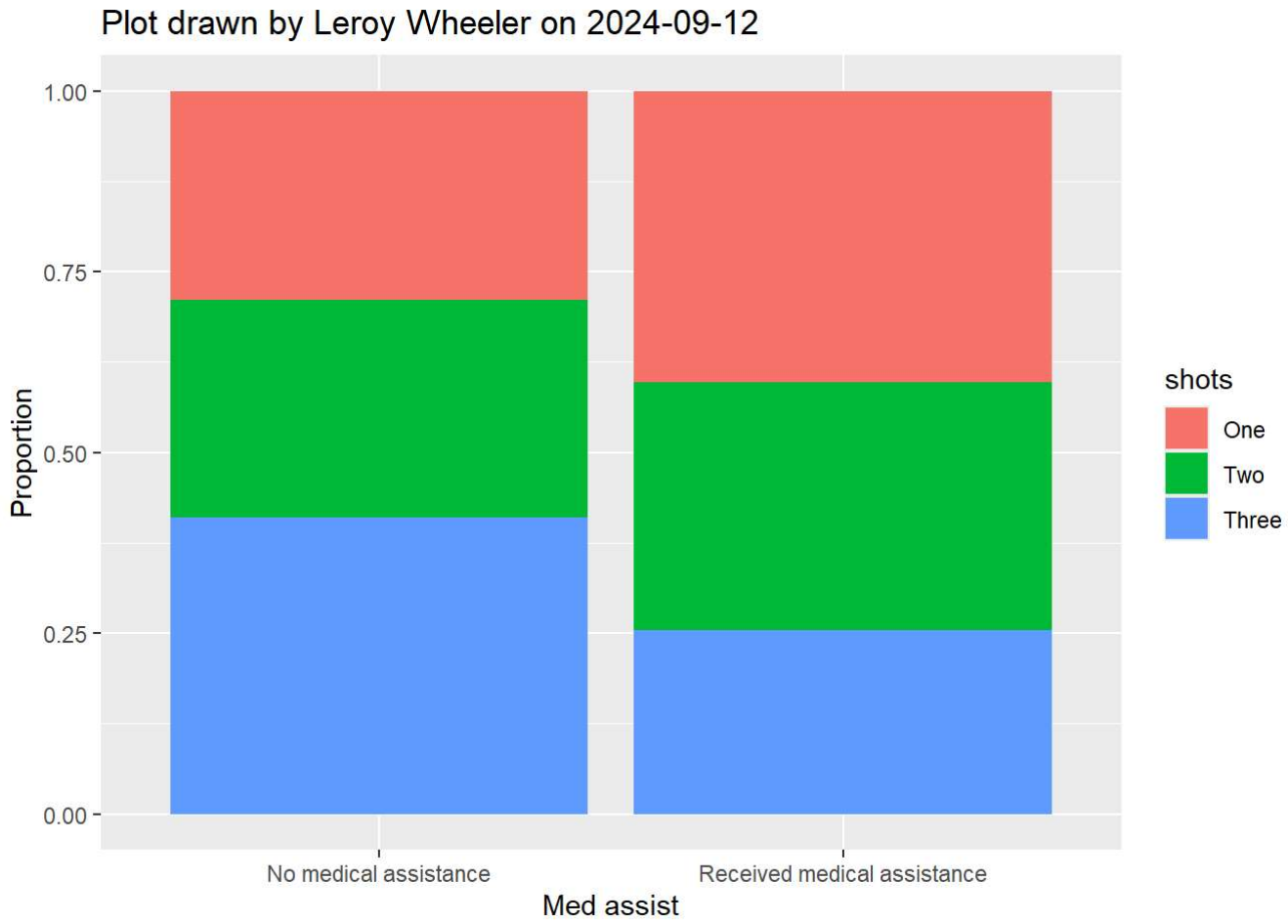
```
        gard |>
          count(medassist, shots) |>
          group_by(medassist) |>
          mutate(row_total=sum(n)) |>
          mutate(pct=round(100*n/row_total))
```

```
# A tibble: 6 × 5
# Groups:   medassist [2]
  medassist                  shots     n row_total   pct
  <fct>                      <fct> <int>     <int> <dbl>
1 No medical assistance      One     329      1138    29
2 No medical assistance      Two     342      1138    30
3 No medical assistance      Three   467      1138    41
4 Received medical assistance One    111       275    40
5 Received medical assistance Two     94       275    34
6 Received medical assistance Three   70       275    25
```

Surprisingly 41% of patients who did not receive medical assistance received all three shots when compared to the 25% of patients who received medical assistance.

Question 8: Draw a bar chart showing the percentages for number of shots received by whether the patient received medical assistance. Interpret this chart.

```
        gard |>
          ggplot(aes(x=medassist, fill=shots)) +
            geom_bar(position="fill") +
            xlab("Med assist") +
            ylab("Proportion") +
            ggtitle("Plot drawn by Leroy Wheeler on 2024-09-12")
```



Patients who did not receive medical assistance were more likely to complete the full round of three Gardisil shots compared to patients who received some medical assistance.