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

Create factors for agegroup

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$agegroup <- factor(
  gard$agegroup,
  levels=0:1,
  labels=c(
    "11 to 17 years",
    "18 to 26 years"))
```

Counts and percentages for agegroup

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

```
# A tibble: 2 × 4
  agegroup           n total   pct
  <fct>          <int> <int> <dbl>
1 11 to 17 years   701  1413    50
2 18 to 26 years   712  1413    50
```

There are roughly the same number of patients 11 to 17 years as there are patients 18 to 26 years.

Question 7

```
gard$medassist <- factor(
  gard$medassist,
  levels=0:1,
  labels=c(
    "patients does not patient have any type of medical assistance",
    "patient have some type of medical assistance"))
```

```
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 patients does not patient have any type of medical assistan…  1138  1413    81
2 patient have some type of medical assistance                   275  1413    19
```

1,138 out of 1,413 total patients (81%). This majority indicates that a significant portion of the sampled population does not receive any medical assistance. 275 out of 1,413 total patients (19%). This smaller percentage reflects the subset of the population that receives some form of medical assistance.

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.

Compare number of shots by age group

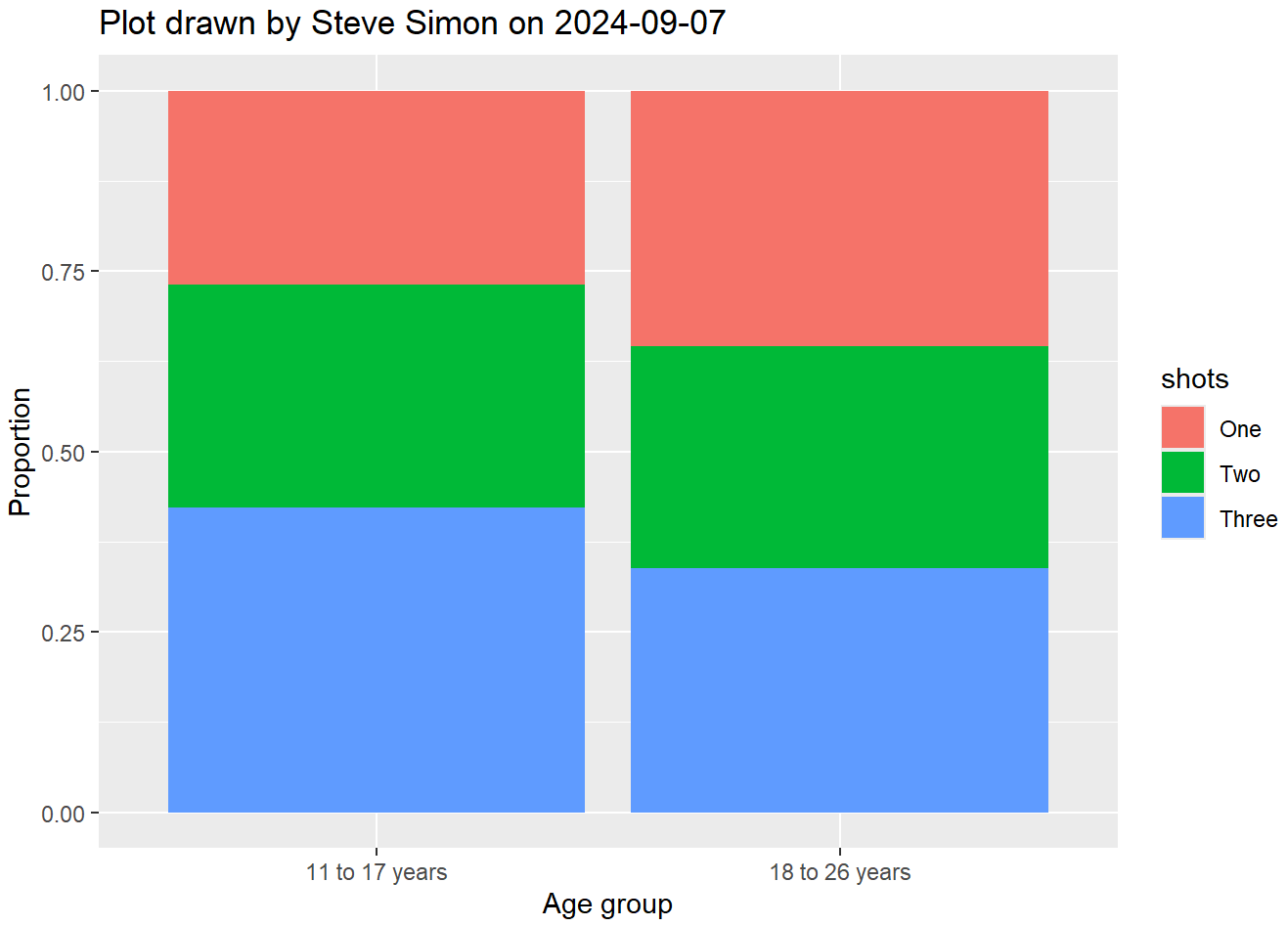
```
gard |>
  count(agegroup, shots) |>
  group_by(agegroup) |>
  mutate(row_total=sum(n)) |>
  mutate(pct=round(100*n/row_total))
```

```
# A tibble: 6 × 5
# Groups:   agegroup [2]
  agegroup       shots     n row_total   pct
  <fct>          <fct> <int>     <int> <dbl>
1 11 to 17 years One     188       701    27
2 11 to 17 years Two     217       701    31
3 11 to 17 years Three   296       701    42
```

```
4 18 to 26 years One     252       712    35
5 18 to 26 years Two     219       712    31
6 18 to 26 years Three   241       712    34
```

Bar chart of shots by age group

```
gard |>
  ggplot(aes(x=agegroup, fill=shots)) +
    geom_bar(position="fill") +
    xlab("Age group") +
    ylab("Proportion") +
    ggtitle("Plot drawn by Steve Simon on 2024-09-07")
```



The probability of getting all three shots was higher in the 11 to 17 year old group compared to the 18 to 26 year old group.

Question 8

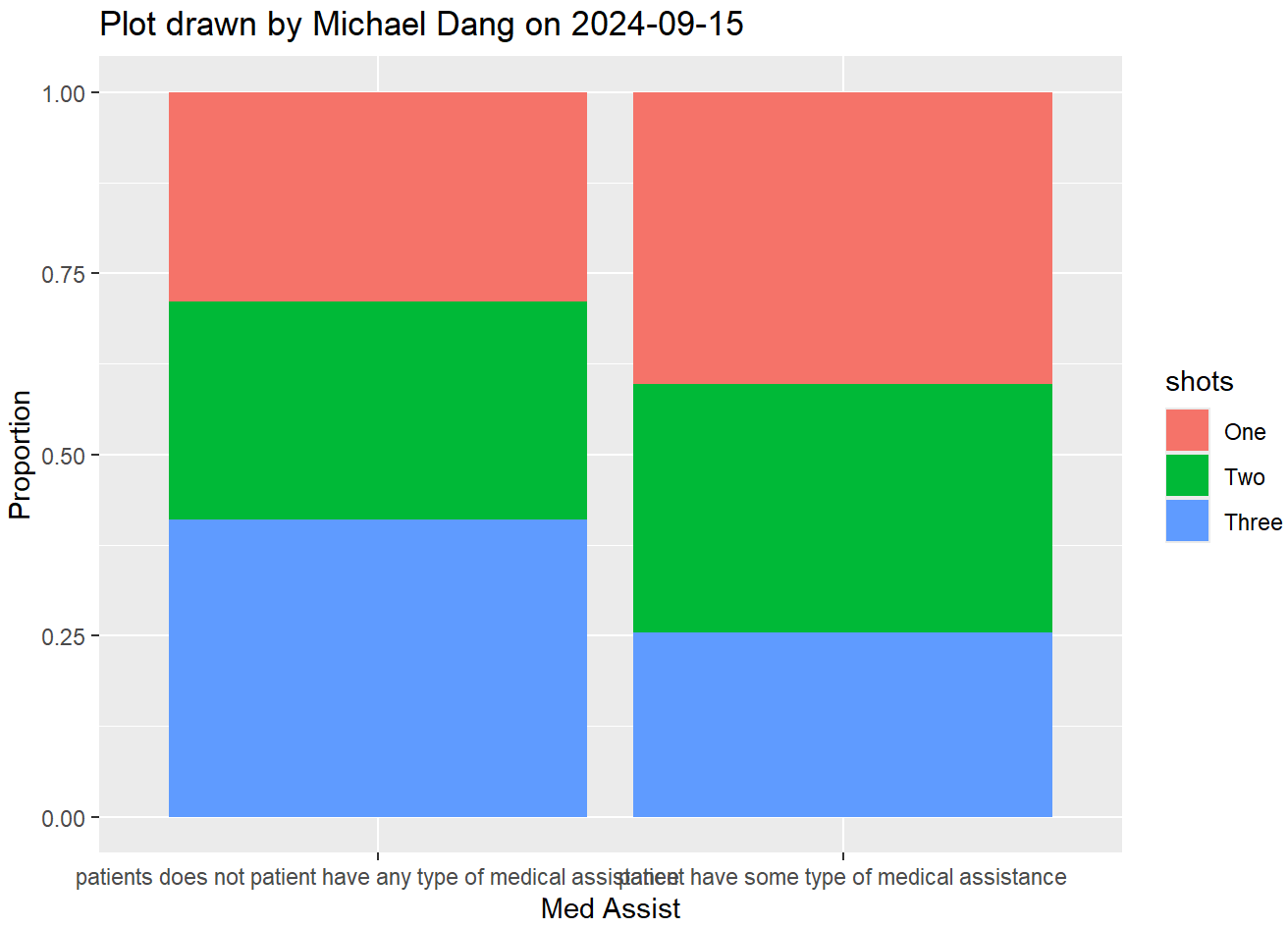
```
gard |>
  count(medassist, shots) |>
  group_by(medassist) |>
```

```
            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 patients does not patient have any type of medica… One     329      1138    29
2 patients does not patient have any type of medica… Two     342      1138    30
3 patients does not patient have any type of medica… Three   467      1138    41
4 patient have some type of medical assistance       One     111       275    40
5 patient have some type of medical assistance       Two      94       275    34
6 patient have some type of medical assistance       Three    70       275    25
```

Almost half (41%) of the patients that did not have any type of medical assistance received all three shots which is higher compared to the (25%) of the patients that had some type of medical assistance.

```
          gard |>
            ggplot(aes(x=medassist, fill=shots)) +
              geom_bar(position="fill") +
              xlab("Med Assist") +
              ylab("Proportion") +
              ggtitle("Plot drawn by Michael Dang on 2024-09-15")
```



The probability of getting all three shots was lower in the patients that have some type of medical assistance compared to the patients that did not have any type of medical assistance.