week6 hip replacement

Melanie Rees-Roberts

2024-11-07

Week 6 practical: Hip replacement data work

#This is work to complete the Hip replacement data practical work for IntroHDS #course, week 6. #The following questions require answering to complete the practical:

#1. Plot 'EQ-5D Index' scores pre and post operation for each gender

#2. Calculate how many patients in this dataset have been told by a doctor that #they have problems caused by a stroke

#3. Create a clean and tidy table with pre and post operation activity levels

Libraries and packages needed

#For this work we require the following packages; tidyverse

#From this package we will use the following libraries

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.3.3
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.3
```

```
## here() starts at C:/Users/mr553/OneDrive - University of Kent/Documents/GitHUB/Week 6 practical/week
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

Import the data

#Data file provided called Hip replacement CCG 1819. Import the data with #read_csv command

```
hip <- read_csv("Hip Replacement CCG 1819.csv")
```

```
## Rows: 28920 Columns: 81
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): Provider Code, Procedure, Year, Age Band, Gender
## dbl (76): Revision Flag, Pre-Op Q Assisted, Pre-Op Q Assisted By, Pre-Op Q S...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Understand, select and tidy the data

```
head(hip)
```

```
## # A tibble: 6 x 81
##   `Provider Code` Procedure       `Revision Flag` Year    `Age Band` Gender
##   <chr>           <chr>                     <dbl> <chr>   <chr>      <chr>
## 1 00C             Hip Replacement               0 2018/19 *          *
## 2 00C             Hip Replacement               0 2018/19 *          *
## 3 00C             Hip Replacement               1 2018/19 *          *
## 4 00C             Hip Replacement               1 2018/19 *          *
## 5 00C             Hip Replacement               0 2018/19 *          *
## 6 00C             Hip Replacement               0 2018/19 *          *
## # i 75 more variables: `Pre-Op Q Assisted` <dbl>, `Pre-Op Q Assisted By` <dbl>,
## #   `Pre-Op Q Symptom Period` <dbl>, `Pre-Op Q Previous Surgery` <dbl>,
## #   `Pre-Op Q Living Arrangements` <dbl>, `Pre-Op Q Disability` <dbl>,
## #   `Heart Disease` <dbl>, `High Bp` <dbl>, Stroke <dbl>, Circulation <dbl>,
## #   `Lung Disease` <dbl>, Diabetes <dbl>, `Kidney Disease` <dbl>,
## #   `Nervous System` <dbl>, `Liver Disease` <dbl>, Cancer <dbl>,
## #   Depression <dbl>, Arthritis <dbl>, `Pre-Op Q Mobility` <dbl>, ...
```

```r
colnames(hip)
```

```
##  [1] "Provider Code"
##  [2] "Procedure"
##  [3] "Revision Flag"
##  [4] "Year"
##  [5] "Age Band"
##  [6] "Gender"
##  [7] "Pre-Op Q Assisted"
##  [8] "Pre-Op Q Assisted By"
##  [9] "Pre-Op Q Symptom Period"
## [10] "Pre-Op Q Previous Surgery"
## [11] "Pre-Op Q Living Arrangements"
## [12] "Pre-Op Q Disability"
## [13] "Heart Disease"
## [14] "High Bp"
## [15] "Stroke"
## [16] "Circulation"
## [17] "Lung Disease"
## [18] "Diabetes"
## [19] "Kidney Disease"
## [20] "Nervous System"
## [21] "Liver Disease"
## [22] "Cancer"
## [23] "Depression"
## [24] "Arthritis"
## [25] "Pre-Op Q Mobility"
## [26] "Pre-Op Q Self-Care"
## [27] "Pre-Op Q Activity"
## [28] "Pre-Op Q Discomfort"
## [29] "Pre-Op Q Anxiety"
## [30] "Pre-Op Q EQ5D Index Profile"
## [31] "Pre-Op Q EQ5D Index"
## [32] "Post-Op Q Assisted"
## [33] "Post-Op Q Assisted By"
## [34] "Post-Op Q Living Arrangements"
## [35] "Post-Op Q Disability"
## [36] "Post-Op Q Mobility"
## [37] "Post-Op Q Self-Care"
## [38] "Post-Op Q Activity"
## [39] "Post-Op Q Discomfort"
## [40] "Post-Op Q Anxiety"
## [41] "Post-Op Q Satisfaction"
## [42] "Post-Op Q Sucess"
## [43] "Post-Op Q Allergy"
## [44] "Post-Op Q Bleeding"
## [45] "Post-Op Q Wound"
## [46] "Post-Op Q Urine"
## [47] "Post-Op Q Further Surgery"
## [48] "Post-Op Q Readmitted"
## [49] "Post-Op Q EQ5D Index Profile"
## [50] "Post-Op Q EQ5D Index"
## [51] "Hip Replacement EQ5D Index Post-Op Q Predicted"
```

```
## [52] "Pre-Op Q EQ VAS"
## [53] "Post-Op Q EQ VAS"
## [54] "Hip Replacement EQ VAS Post-Op Q Predicted"
## [55] "Hip Replacement Pre-Op Q Pain"
## [56] "Hip Replacement Pre-Op Q Sudden Pain"
## [57] "Hip Replacement Pre-Op Q Night Pain"
## [58] "Hip Replacement Pre-Op Q Washing"
## [59] "Hip Replacement Pre-Op Q Transport"
## [60] "Hip Replacement Pre-Op Q Dressing"
## [61] "Hip Replacement Pre-Op Q Shopping"
## [62] "Hip Replacement Pre-Op Q Walking"
## [63] "Hip Replacement Pre-Op Q Limping"
## [64] "Hip Replacement Pre-Op Q Stairs"
## [65] "Hip Replacement Pre-Op Q Standing"
## [66] "Hip Replacement Pre-Op Q Work"
## [67] "Hip Replacement Pre-Op Q Score"
## [68] "Hip Replacement Post-Op Q Pain"
## [69] "Hip Replacement Post-Op Q Sudden Pain"
## [70] "Hip Replacement Post-Op Q Night Pain"
## [71] "Hip Replacement Post-Op Q Washing"
## [72] "Hip Replacement Post-Op Q Transport"
## [73] "Hip Replacement Post-Op Q Dressing"
## [74] "Hip Replacement Post-Op Q Shopping"
## [75] "Hip Replacement Post-Op Q Walking"
## [76] "Hip Replacement Post-Op Q Limping"
## [77] "Hip Replacement Post-Op Q Stairs"
## [78] "Hip Replacement Post-Op Q Standing"
## [79] "Hip Replacement Post-Op Q Work"
## [80] "Hip Replacement Post-Op Q Score"
## [81] "Hip Replacement OHS Post-Op Q Predicted"
```

```
#Check this is for a single year - then dont need the year variable
hip %>%
count(Year)
```

```
## # A tibble: 1 x 2
##   Year        n
##   <chr>   <int>
## 1 2018/19 28920
```

#For this exercise, we need the following data variables: #Gender, Strike, Pre-OP Q EQ5D Index, Post-Op Q EQ5D Index, Pre-Op Q Activity, Post-Op Q Activity #Create tibble with required variables, shortening the variable names where appropriate

```
hip2 <- hip %>%
  select('Gender','Pre-Op Q EQ5D Index','Post-Op Q EQ5D Index','Stroke','Pre-Op Q Activity','Post-Op Q A
  rename(EQ5D_pre = 'Pre-Op Q EQ5D Index',
         EQ5D_post = 'Post-Op Q EQ5D Index',
         activity_pre = 'Pre-Op Q Activity',
         activity_post = 'Post-Op Q Activity')

head(hip2)
```

```
## # A tibble: 6 x 6
##   Gender EQ5D_pre EQ5D_post Stroke activity_pre activity_post
##   <chr>     <dbl>     <dbl>  <dbl>        <dbl>         <dbl>
## 1 *        NA         0.516      9            9             2
## 2 *        -0.003    NA          9            3             9
## 3 *        NA        -0.074      9            9             3
## 4 *         0.03      0.796      9            3             1
## 5 *        -0.239     0.62       9            3             2
## 6 *         0.159     0.691      9            2             2
```

#Remove missing values - create this into new data frame hip3

```
hip2$Gender %>% table()
```

```
## .
##     *     1     2
##  2309 10255 16356
```

```
hip3 <- hip2 %>%
  drop_na() %>%
  filter(Gender !='*')

head(hip3)
```

```
## # A tibble: 6 x 6
##   Gender EQ5D_pre EQ5D_post Stroke activity_pre activity_post
##   <chr>     <dbl>     <dbl>  <dbl>        <dbl>         <dbl>
## 1 1        -0.016     0.516      9            2             2
## 2 1         0.159     0.743      9            2             2
## 3 1         0.03      0.727      9            3             1
## 4 1         0.587     0.85       9            2             1
## 5 1         0.623     0.796      9            1             1
## 6 1         0.691     1          9            2             1
```

##Tidy the data - move pre and post into a time variable and th EQ5D and Activity into variables #Saved into a new dataset tidy_hip

```
tidy_hip <- hip3 %>%
  pivot_longer(c(EQ5D_pre,EQ5D_post),
               names_to = 'Time_EQ5D',
               names_prefix = 'EQ5D_',
               values_to = 'EQ5D')


tidy_hip <- tidy_hip %>%
  pivot_longer(c(activity_pre, activity_post),
               names_to = 'Time_activity',
               names_prefix = 'activity_',
               values_to = 'activity')


head(tidy_hip)
```

```
## # A tibble: 6 x 6
##   Gender Stroke Time_EQ5D   EQ5D Time_activity activity
##   <chr>   <dbl> <chr>      <dbl> <chr>            <dbl>
## 1 1           9 pre       -0.016 pre                  2
## 2 1           9 pre       -0.016 post                 2
## 3 1           9 post       0.516 pre                  2
## 4 1           9 post       0.516 post                 2
## 5 1           9 pre        0.159 pre                  2
## 6 1           9 pre        0.159 post                 2
```

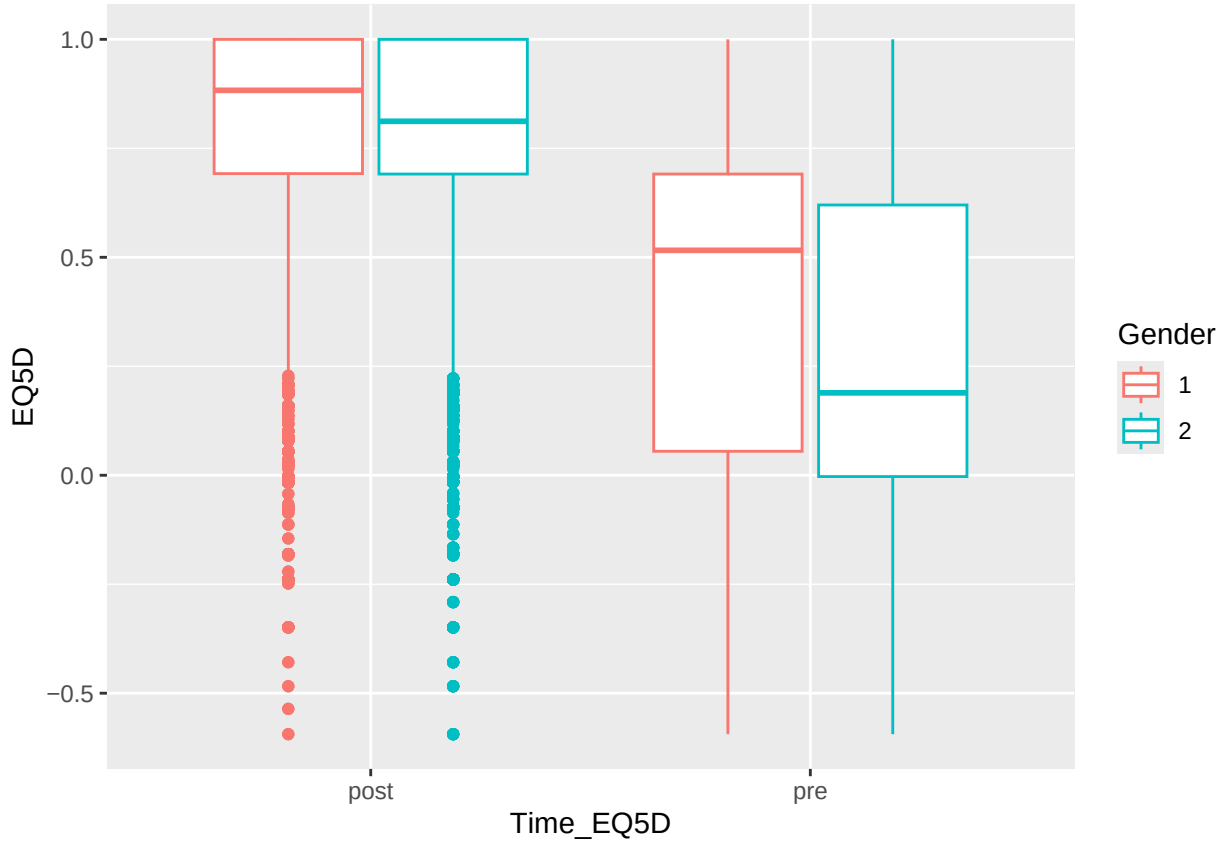
##Answer the questions

##1. Plot 'EQ-5D Index' scores pre and post operation for each gender

```
#From tidy_hip we can now select the data and plot the scores

tidy_hip_plot <- tidy_hip %>%
  select(EQ5D, Time_EQ5D,Gender
         )

head(tidy_hip_plot)
```

```
## # A tibble: 6 x 3
##     EQ5D Time_EQ5D Gender
##    <dbl> <chr>     <chr>
## 1 -0.016 pre       1
## 2 -0.016 pre       1
## 3  0.516 post      1
## 4  0.516 post      1
## 5  0.159 pre       1
## 6  0.159 pre       1
```

```
#Now plot the data pre and post for each gender

tidy_hip_plot %>%
  ggplot() +
  geom_boxplot(aes(x=Time_EQ5D, y= EQ5D, colour = Gender))
```



##2. Calculate how many patients in this dataset have been told by a doctor that #they have problems caused by a stroke

```
#We need to return to our table where we did not make the data tidy - this may cause duplication of Str
#calculate the how many patients have been told they have problems caused by stroke

hip2$Stroke %>% table()
```

```
## .
##     1     9
##   400 28520
```

A total of 400 patients in this dataset have been told that they have problems caused by stroke.

##3. Create a clean and tidy table with pre and post operation activity levels #We converted the pre- and post- activity levels earlier in the file - see code above. However, to show just these variables

```
head(tidy_hip)
```

```
## # A tibble: 6 x 6
##   Gender Stroke Time_EQ5D   EQ5D Time_activity activity
##   <chr>   <dbl> <chr>      <dbl> <chr>            <dbl>
```

```
## 1 1           9 pre       -0.016 pre                 2
## 2 1           9 pre       -0.016 post                2
## 3 1           9 post       0.516 pre                 2
## 4 1           9 post       0.516 post                2
## 5 1           9 pre        0.159 pre                 2
## 6 1           9 pre        0.159 post                2
```

```
tidy_hip_activity <- tidy_hip %>%
  select(Time_activity,activity)

head(tidy_hip_activity)
```

```
## # A tibble: 6 x 2
##   Time_activity activity
##   <chr>            <dbl>
## 1 pre                  2
## 2 post                 2
## 3 pre                  2
## 4 post                 2
## 5 pre                  2
## 6 post                 2
```