

hip_replacement_operations

Uyen Nguyen

25/11/2025

Aim

1. Plot 'EQ-5D Index' scores (a combination of five key criteria concerning patients' self-reported general health) pre and post a hip replacement 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

Load packages

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyrr    1.3.1
## v purrr    1.2.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Read in data

```
hip_data <- 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.
```

Prepare the data

This includes three steps: inspecting the data, selecting only the variables we want, and dealing with missing values.

```
glimpse(hip_data)
```

```

## Rows: 28,920
## Columns: 81
## $ 'Provider Code'
## $ Procedure
## $ 'Revision Flag'
## $ Year
## $ 'Age Band'
## $ Gender
## $ 'Pre-Op Q Assisted'
## $ 'Pre-Op Q Assisted By'
## $ 'Pre-Op Q Symptom Period'
## $ 'Pre-Op Q Previous Surgery'
## $ 'Pre-Op Q Living Arrangements'
## $ 'Pre-Op Q Disability'
## $ 'Heart Disease'
## $ 'High Bp'
## $ Stroke
## $ Circulation
## $ 'Lung Disease'
## $ Diabetes
## $ 'Kidney Disease'
## $ 'Nervous System'
## $ 'Liver Disease'
## $ Cancer
## $ Depression
## $ Arthritis
## $ 'Pre-Op Q Mobility'
## $ 'Pre-Op Q Self-Care'
## $ 'Pre-Op Q Activity'
## $ 'Pre-Op Q Discomfort'
## $ 'Pre-Op Q Anxiety'
## $ 'Pre-Op Q EQ5D Index Profile'
## $ 'Pre-Op Q EQ5D Index'
## $ 'Post-Op Q Assisted'
## $ 'Post-Op Q Assisted By'
## $ 'Post-Op Q Living Arrangements'
## $ 'Post-Op Q Disability'
## $ 'Post-Op Q Mobility'
## $ 'Post-Op Q Self-Care'
## $ 'Post-Op Q Activity'
## $ 'Post-Op Q Discomfort'
## $ 'Post-Op Q Anxiety'
## $ 'Post-Op Q Satisfaction'
## $ 'Post-Op Q Success'
## $ 'Post-Op Q Allergy'
## $ 'Post-Op Q Bleeding'
## $ 'Post-Op Q Wound'
<chr> "00C", "00C", "00C", ~
<chr> "Hip Replacement", "H~
<dbl> 0, 0, 1, 1, 0, 0, 0, ~
<chr> "2018/19", "2018/19", ~
<chr> "*", "*", "*", "*", "~
<chr> "*", "*", "*", "*", "~
<dbl> 2, 2, 1, 2, 2, 2, 2, ~
<dbl> 0, 0, 0, 0, 0, 0, 0, ~
<dbl> 4, 2, 4, 1, 2, 1, 1, ~
<dbl> 2, 1, 1, 1, 2, 2, 1, ~
<dbl> 1, 1, 2, 2, 1, 2, 1, ~
<dbl> 9, 1, 1, 1, 2, 1, 2, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 1, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 1, ~
<dbl> 9, 9, 9, 9, 1, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 1, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 1, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 9, ~
<dbl> 9, 9, 9, 9, 9, 9, 1, ~
<dbl> 9, 9, 9, 9, 9, 9, 1, ~
<dbl> 9, 9, 9, 9, 1, 9, 9, ~
<dbl> 9, 1, 1, 1, 1, 1, 1, ~
<dbl> 2, 2, 9, 2, 2, 2, 2, ~
<dbl> 1, 2, 9, 1, 2, 1, 1, ~
<dbl> 9, 3, 9, 3, 3, 2, 2, ~
<dbl> 9, 3, 9, 3, 3, 3, 2, ~
<dbl> 9, 1, 9, 2, 3, 1, 1, ~
<dbl> 21999, 22331, 99999, ~
<dbl> NA, -0.003, NA, 0.030~
<dbl> 2, 2, 1, 2, 2, 2, 1, ~
<dbl> 9, 9, 1, 9, 9, 9, 1, ~
<dbl> 1, 1, 2, 2, 1, 2, 1, ~
<dbl> 2, 9, 1, 2, 1, 2, 2, ~
<dbl> 2, 9, 2, 1, 2, 2, 1, ~
<dbl> 2, 1, 2, 1, 1, 1, 1, ~
<dbl> 2, 9, 3, 1, 2, 2, 1, ~
<dbl> 2, 1, 3, 2, 2, 2, 1, ~
<dbl> 2, 1, 2, 1, 2, 1, 1, ~
<dbl> 2, 3, 2, 1, 3, 1, 1, ~
<dbl> 1, 1, 1, 1, 2, 2, 1, ~
<dbl> 2, 2, 2, 2, 2, 9, 9, ~
<dbl> 2, 2, 2, 2, 2, 9, 9, ~
<dbl> 2, 2, 1, 2, 2, 9, 9, ~

```

```

## $ 'Post-Op Q Urine'
## $ 'Post-Op Q Further Surgery'
## $ 'Post-Op Q Readmitted'
## $ 'Post-Op Q EQ5D Index Profile'
## $ 'Post-Op Q EQ5D Index'
## $ 'Hip Replacement EQ5D Index Post-Op Q Predicted'
## $ 'Pre-Op Q EQ VAS'
## $ 'Post-Op Q EQ VAS'
## $ 'Hip Replacement EQ VAS Post-Op Q Predicted'
## $ 'Hip Replacement Pre-Op Q Pain'
## $ 'Hip Replacement Pre-Op Q Sudden Pain'
## $ 'Hip Replacement Pre-Op Q Night Pain'
## $ 'Hip Replacement Pre-Op Q Washing'
## $ 'Hip Replacement Pre-Op Q Transport'
## $ 'Hip Replacement Pre-Op Q Dressing'
## $ 'Hip Replacement Pre-Op Q Shopping'
## $ 'Hip Replacement Pre-Op Q Walking'
## $ 'Hip Replacement Pre-Op Q Limping'
## $ 'Hip Replacement Pre-Op Q Stairs'
## $ 'Hip Replacement Pre-Op Q Standing'
## $ 'Hip Replacement Pre-Op Q Work'
## $ 'Hip Replacement Pre-Op Q Score'
## $ 'Hip Replacement Post-Op Q Pain'
## $ 'Hip Replacement Post-Op Q Sudden Pain'
## $ 'Hip Replacement Post-Op Q Night Pain'
## $ 'Hip Replacement Post-Op Q Washing'
## $ 'Hip Replacement Post-Op Q Transport'
## $ 'Hip Replacement Post-Op Q Dressing'
## $ 'Hip Replacement Post-Op Q Shopping'
## $ 'Hip Replacement Post-Op Q Walking'
## $ 'Hip Replacement Post-Op Q Limping'
## $ 'Hip Replacement Post-Op Q Stairs'
## $ 'Hip Replacement Post-Op Q Standing'
## $ 'Hip Replacement Post-Op Q Work'
## $ 'Hip Replacement Post-Op Q Score'
## $ 'Hip Replacement OHS Post-Op Q Predicted'
<dbl> 2, 2, 2, 2, 2, 1, 9, ~
<dbl> 2, 2, 1, 2, 2, 2, 2, ~
<dbl> 2, 2, 1, 2, 2, 2, 2, ~
<dbl> 22222, 91911, 22332, ~
<dbl> 0.516, NA, -0.074, 0.~
<dbl> NA, NA, NA, 0.5154424~
<dbl> 999, 999, 999, 50, 30~
<dbl> 70, 999, 80, 90, 70, ~
<dbl> NA, NA, NA, 60.05266, ~
<dbl> 1, 0, 0, 0, 0, 0, 1, ~
<dbl> 0, 1, 0, 0, 0, 1, 4, ~
<dbl> 2, 0, 1, 0, 0, 1, 1, ~
<dbl> 3, 1, 1, 2, 2, 4, 4, ~
<dbl> 2, 1, 1, 0, 1, 2, 2, ~
<dbl> 1, 0, 1, 0, 1, 4, 2, ~
<dbl> 3, 2, 0, 0, 0, 0, 3, ~
<dbl> 2, 0, 1, 1, 1, 3, 3, ~
<dbl> 2, 0, 0, 1, 0, 0, 0, ~
<dbl> 2, 1, 1, 1, 1, 2, 4, ~
<dbl> 1, 1, 1, 2, 1, 1, 4, ~
<dbl> 1, 1, 0, 1, 0, 0, 4, ~
<dbl> 20, 8, 7, 8, 7, 18, 3~
<dbl> 3, 4, 2, 2, 4, 2, 2, ~
<dbl> 4, 4, 4, 2, 2, 2, 4, ~
<dbl> 4, 4, 4, 1, 4, 2, 4, ~
<dbl> 4, 3, 3, 4, 3, 4, 4, ~
<dbl> 4, 4, 2, 3, 3, 2, 4, ~
<dbl> 2, 4, 3, 3, 4, 4, 3, ~
<dbl> 4, 2, 0, 3, 2, 0, 4, ~
<dbl> 4, 3, 1, 4, 3, 2, 4, ~
<dbl> 3, 1, 1, 4, 2, 0, 3, ~
<dbl> 4, 1, 1, 3, 2, 4, 4, ~
<dbl> 3, 4, 3, 3, 4, 2, 4, ~
<dbl> 4, 4, 2, 4, 2, 2, 3, ~
<dbl> 43, 38, 26, 36, 35, 2~
<dbl> 42.20017, 35.29577, 2~

```

Select gender and quality of life score pre and post operation

```

gender_EQ5D <- hip_data %>%
  select(`Gender`, `Pre-Op Q EQ5D Index`, `Post-Op Q EQ5D Index`) %>%
  rename(
    EQ5D_Pre = `Pre-Op Q EQ5D Index`,
    EQ5D_Post = `Post-Op Q EQ5D Index`
  )

```

```
head(gender_EQ5D)
```

```

## # A tibble: 6 x 3
##   Gender EQ5D_Pre EQ5D_Post
##   <chr>     <dbl>      <dbl>
## 1 *         NA          0.516

```

```

## 2 *      -0.003    NA
## 3 *      NA        -0.074
## 4 *      0.03      0.796
## 5 *      -0.239    0.62
## 6 *      0.159    0.691

```

Identify and remove missing values

```
gender_EQ5D$Gender %>% unique()
```

```
## [1] "*" "1" "2"
```

```
gender_EQ5D$Gender %>% table()
```

```

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

```

```
gender_EQ5D %>% summary()
```

```

##      Gender          EQ5D_Pre       EQ5D_Post
## Length:28920      Min. :-0.5940      Min. :-0.5940
## Class :character  1st Qu.: 0.0300    1st Qu.: 0.6910
## Mode  :character  Median : 0.3640    Median : 0.8150
##                  Mean  : 0.3357    Mean  : 0.7975
##                  3rd Qu.: 0.6200    3rd Qu.: 1.0000
##                  Max.  : 1.0000    Max.  : 1.0000
##                  NA's   :1794      NA's   :1104

```

```
gender_EQ5D_noNA <- gender_EQ5D %>%
```

```
  drop_na() %>%
  filter(Gender != '*')
```

```
table(gender_EQ5D_noNA$Gender)
```

```

## 
##   1     2
## 9381 14661

```

```
gender_EQ5D_noNA <- gender_EQ5D_noNA %>%
```

```
  mutate(Gender = recode(Gender,
                         `1` = "Male",
                         `2` = "Female"))
```

```
summary(gender_EQ5D_noNA)
```

```

##      Gender          EQ5D_Pre       EQ5D_Post
## Length:24042      Min. :-0.594      Min. :-0.5940
## Class :character  1st Qu.: 0.055    1st Qu.: 0.6910
## Mode  :character  Median : 0.516    Median : 0.8150
##                  Mean  : 0.339    Mean  : 0.7995
##                  3rd Qu.: 0.656    3rd Qu.: 1.0000
##                  Max.  : 1.000    Max.  : 1.0000

```

Check that data is tidy

The data frame is not tidy, because the column names EQ5D_Pre and EQ5D_Post contain *data*: the time point when EQ5D was measured: pre or post operation.

```
head(gender_EQ5D_noNA)

## # A tibble: 6 x 3
##   Gender EQ5D_Pre EQ5D_Post
##   <chr>     <dbl>     <dbl>
## 1 Male      -0.016     0.516
## 2 Male       0.159     0.743
## 3 Male       0.03      0.727
## 4 Male       0.587     0.85
## 5 Male       0.623     0.796
## 6 Male       0.691     1
```

```
tidy_gender_EQ5D_noNA <- gender_EQ5D_noNA %>%
  pivot_longer(c(EQ5D_Pre,EQ5D_Post),
               names_to = 'Time',    # the name of the column to create from the data stored in the original column
               names_prefix = 'EQ5D_', # remove this text from the start of each variable name
               values_to = 'EQ5D' # the name of the column to create from the data stored in cell values
  )
```

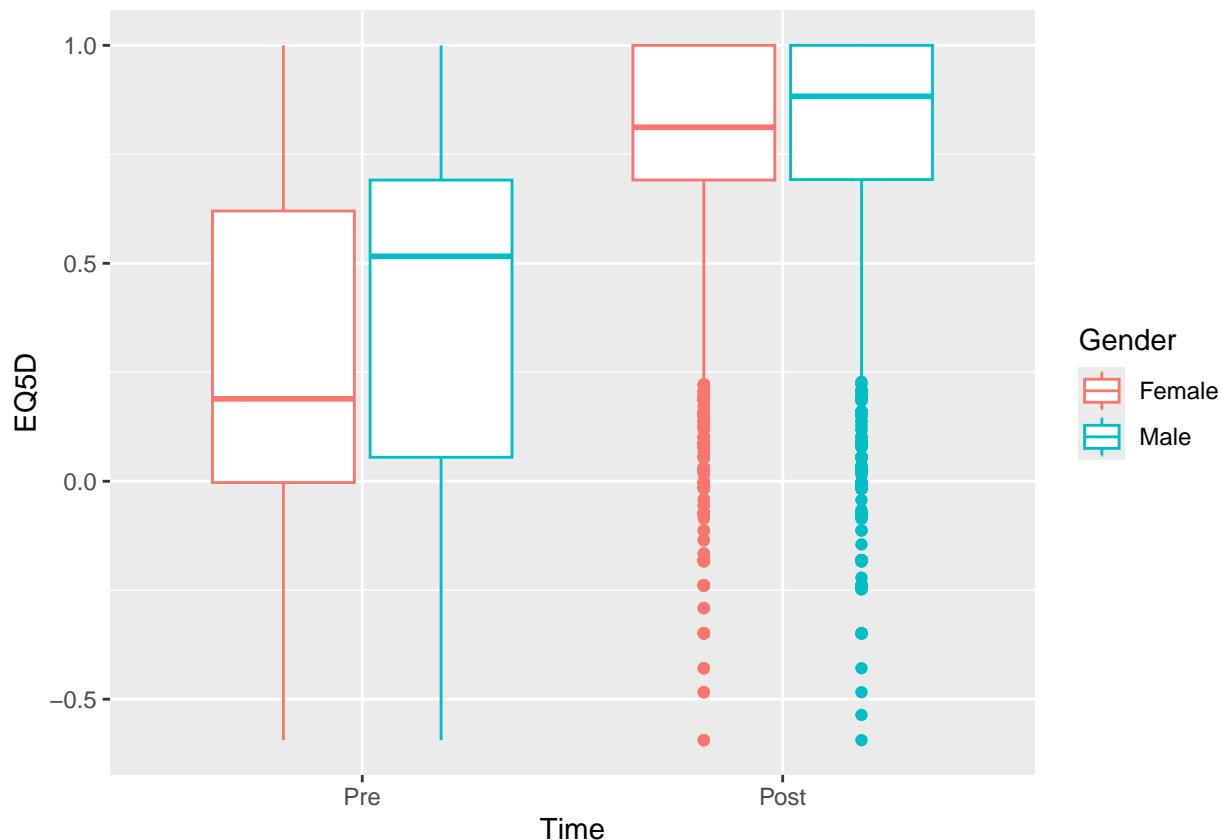
```
head(tidy_gender_EQ5D_noNA)
```

```
## # A tibble: 6 x 3
##   Gender Time     EQ5D
##   <chr>  <chr>   <dbl>
## 1 Male   Pre     -0.016
## 2 Male   Post    0.516
## 3 Male   Pre     0.159
## 4 Male   Post    0.743
## 5 Male   Pre     0.03
## 6 Male   Post    0.727
```

Plot quality of life pre and post operation for each gender

```
# Turn Time into a "factor" so we can order the categories any way we want
# otherwise they are alphabetical and "Post" ends up before "Pre"
tidy_gender_EQ5D_noNA$Time <- factor(tidy_gender_EQ5D_noNA$Time,levels=c('Pre','Post'))

# ggplot creates a blank canvas, to which we add a boxplot with "geom_boxplot"
tidy_gender_EQ5D_noNA %>%
  ggplot() +
  geom_boxplot(aes(x = Time, y = EQ5D, colour = Gender))
```



Number of patients in this dataset have been told by a doctor that they have problems caused by a stroke

```
patient_stroke <- hip_data %>%
  filter(Stroke == "1") %>%
  summarise(n=n())

paste("Number of patients:", patient_stroke %>% pull(n))

## [1] "Number of patients: 400"
```

Table of pre and post operation activity levels

Select activity pre and post operation

```
activity <- hip_data %>%
  select(`Pre-Op Q Activity`, `Post-Op Q Activity`) %>%
  rename(
    Activity_Pre = `Pre-Op Q Activity`,
    Activity_Post = `Post-Op Q Activity`
  )
```

```
head(activity)

## # A tibble: 6 x 2
##   Activity_Pre Activity_Post
##       <dbl>        <dbl>
## 1         9          2
## 2         3          9
## 3         9          3
## 4         3          1
## 5         3          2
## 6         2          2
```

Remove missing value

```
activity %>% table()
```

```
##           Activity_Post
## Activity_Pre      1     2     3     9
##   1     1343    250    14    18
##   2    12393   7513   335   381
##   3    2196   2714   476   113
##   9     670    441    33    30
```

```
activity %>% summary()
```

```
##   Activity_Pre   Activity_Post
##   Min.   :1.000   Min.   :1.000
##   1st Qu.:2.000   1st Qu.:1.000
##   Median :2.000   Median :1.000
##   Mean   :2.418   Mean   :1.587
##   3rd Qu.:2.000   3rd Qu.:2.000
##   Max.   :9.000   Max.   :9.000
```

```
activity_noNA <- activity %>%
  drop_na() %>%
  filter(Activity_Pre != 9,
         Activity_Post != 9)
```

```
activity_noNA %>% table()
```

```
##           Activity_Post
## Activity_Pre      1     2     3
##   1     1343    250    14
##   2    12393   7513   335
##   3    2196   2714   476
```

```
summary(activity_noNA)
```

```
##   Activity_Pre   Activity_Post
##   Min.    :1.000   Min.    :1.000
##   1st Qu.:2.000   1st Qu.:1.000
##   Median  :2.000   Median  :1.000
##   Mean    :2.139   Mean    :1.445
##   3rd Qu.:2.000   3rd Qu.:2.000
##   Max.    :3.000   Max.    :3.000
```

Check data is tidy

```
head(activity_noNA)
```

```
## # A tibble: 6 x 2
##   Activity_Pre Activity_Post
##       <dbl>        <dbl>
## 1         3          1
## 2         3          2
## 3         2          2
## 4         2          1
## 5         2          1
## 6         2          1
```

```
tidy_activity_noNA <- activity_noNA %>%
  pivot_longer(c(Activity_Pre,Activity_Post),
               names_to = 'Time',
               names_prefix = 'Activity_',
               values_to = 'Activity'
  )
```

```
head(tidy_activity_noNA)
```

```
## # A tibble: 6 x 2
##   Time   Activity
##   <chr>    <dbl>
## 1 Pre      3
## 2 Post     1
## 3 Pre      3
## 4 Post     2
## 5 Pre      2
## 6 Post     2
```