Hip Replacement Exercise week 6

Emma Cliffe

2025-10-26

Aim

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

Load packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.3

## Warning: package 'ggplot2' was built under R version 4.4.3

## Warning: package 'tibble' was built under R version 4.4.3

## Warning: package 'tidyr' was built under R version 4.4.3

## Warning: package 'readr' was built under R version 4.4.3

## Warning: package 'purrr' was built under R version 4.4.3

## Warning: package 'dplyr' was built under R version 4.4.3

## Warning: package 'stringr' was built under R version 4.4.3

## Warning: package 'forcats' was built under R version 4.4.3

## Warning: package 'lubridate' was built under R version 4.4.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- 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
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.4.3

## here() starts at C:/Users/ECliffe ABDN/OneDrive/Documents/IntroHDS/GitHub/Intro2hdsR
```

Read in data

```
hip_data <- read_csv(here("./Inputs/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.
```

```
head(hip_data)
```

```
## # 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>, ...
```

Prepare the data

Inspect

```
glimpse(hip_data)
```

```
## Rows: 28,920
## Columns: 81
## $ `Provider Code`                                  <chr> "00C", "00C", "00C", ~
## $ Procedure                                        <chr> "Hip Replacement", "H~
## $ ‘Revision Flag‘                                  <dbl> 0, 0, 1, 1, 0, 0, 0, ~
## $ Year                                             <chr> "2018/19", "2018/19",~
## $ ‘Age Band‘                                       <chr> "*", "*", "*", "*", "~
## $ Gender                                           <chr> "*", "*", "*", "*", "~
## $ ‘Pre-Op Q Assisted‘                              <dbl> 2, 2, 1, 2, 2, 2, 2, ~
## $ ‘Pre-Op Q Assisted By‘                           <dbl> 0, 0, 0, 0, 0, 0, 0, ~
## $ ‘Pre-Op Q Symptom Period‘                        <dbl> 4, 2, 4, 1, 2, 1, 1, ~
## $ ‘Pre-Op Q Previous Surgery‘                      <dbl> 2, 1, 1, 1, 2, 2, 1, ~
## $ ‘Pre-Op Q Living Arrangements‘                   <dbl> 1, 1, 2, 2, 1, 2, 1, ~
## $ ‘Pre-Op Q Disability‘                            <dbl> 9, 1, 1, 1, 2, 1, 2, ~
## $ ‘Heart Disease‘                                  <dbl> 9, 9, 9, 9, 9, 9, 9, ~
## $ ‘High Bp‘                                        <dbl> 9, 9, 9, 9, 9, 1, 9, ~
## $ Stroke                                           <dbl> 9, 9, 9, 9, 9, 9, 1, ~
## $ Circulation                                      <dbl> 9, 9, 9, 9, 1, 9, 9, ~
## $ ‘Lung Disease‘                                   <dbl> 9, 9, 9, 9, 9, 9, 9, ~
## $ Diabetes                                         <dbl> 9, 9, 9, 9, 9, 9, 9, ~
## $ ‘Kidney Disease‘                                 <dbl> 9, 9, 9, 9, 9, 1, 9, ~
## $ ‘Nervous System‘                                 <dbl> 9, 9, 9, 9, 9, 9, 9, ~
## $ ‘Liver Disease‘                                  <dbl> 9, 9, 9, 9, 9, 9, 1, ~
## $ Cancer                                           <dbl> 9, 9, 9, 9, 9, 9, 1, ~
## $ Depression                                       <dbl> 9, 9, 9, 1, 9, 9, 9, ~
## $ Arthritis                                        <dbl> 9, 1, 1, 1, 1, 1, 9, ~
## $ ‘Pre-Op Q Mobility‘                              <dbl> 2, 2, 9, 2, 2, 2, 2, ~
## $ ‘Pre-Op Q Self-Care‘                             <dbl> 1, 2, 9, 1, 2, 1, 1, ~
## $ ‘Pre-Op Q Activity‘                              <dbl> 9, 3, 9, 3, 3, 2, 2, ~
## $ ‘Pre-Op Q Discomfort‘                            <dbl> 9, 3, 9, 3, 3, 3, 2, ~
## $ ‘Pre-Op Q Anxiety‘                               <dbl> 9, 1, 9, 2, 3, 1, 1, ~
## $ ‘Pre-Op Q EQ5D Index Profile‘                    <dbl> 21999, 22331, 99999, ~
## $ ‘Pre-Op Q EQ5D Index‘                            <dbl> NA, -0.003, NA, 0.030~
## $ ‘Post-Op Q Assisted‘                             <dbl> 2, 2, 1, 2, 2, 2, 1, ~
## $ ‘Post-Op Q Assisted By‘                          <dbl> 9, 9, 1, 9, 9, 9, 1, ~
## $ ‘Post-Op Q Living Arrangements‘                  <dbl> 1, 1, 2, 2, 1, 2, 1, ~
## $ ‘Post-Op Q Disability‘                           <dbl> 2, 9, 1, 2, 1, 2, 2, ~
## $ ‘Post-Op Q Mobility‘                             <dbl> 2, 9, 2, 1, 2, 2, 1, ~
## $ ‘Post-Op Q Self-Care‘                            <dbl> 2, 1, 2, 1, 1, 1, 1, ~
## $ ‘Post-Op Q Activity‘                             <dbl> 2, 9, 3, 1, 2, 2, 1, ~
## $ ‘Post-Op Q Discomfort‘                           <dbl> 2, 1, 3, 2, 2, 2, 1, ~
## $ ‘Post-Op Q Anxiety‘                              <dbl> 2, 1, 2, 1, 2, 1, 1, ~
## $ ‘Post-Op Q Satisfaction‘                         <dbl> 2, 3, 2, 1, 3, 1, 1, ~
## $ ‘Post-Op Q Sucess‘                               <dbl> 1, 1, 1, 1, 2, 2, 1, ~
## $ ‘Post-Op Q Allergy‘                              <dbl> 2, 2, 2, 2, 2, 9, 9, ~
## $ ‘Post-Op Q Bleeding‘                             <dbl> 2, 2, 2, 2, 2, 9, 9, ~
## $ ‘Post-Op Q Wound‘                                <dbl> 2, 2, 1, 2, 2, 9, 9, ~
## $ ‘Post-Op Q Urine‘                                <dbl> 2, 2, 2, 2, 2, 1, 9, ~
## $ ‘Post-Op Q Further Surgery‘                      <dbl> 2, 2, 1, 2, 2, 2, 2, ~
## $ ‘Post-Op Q Readmitted‘                           <dbl> 2, 2, 1, 2, 2, 2, 2, ~
## $ ‘Post-Op Q EQ5D Index Profile‘                   <dbl> 22222, 91911, 22332, ~
## $ ‘Post-Op Q EQ5D Index‘                           <dbl> 0.516, NA, -0.074, 0.~
## $ ‘Hip Replacement EQ5D Index Post-Op Q Predicted‘ <dbl> NA, NA, NA, 0.5154424~
## $ ‘Pre-Op Q EQ VAS‘                                <dbl> 999, 999, 999, 50, 30~
## $ ‘Post-Op Q EQ VAS‘                               <dbl> 70, 999, 80, 90, 70, ~
## $ ‘Hip Replacement EQ VAS Post-Op Q Predicted‘     <dbl> NA, NA, NA, 60.05266,~
## $ ‘Hip Replacement Pre-Op Q Pain‘                  <dbl> 1, 0, 0, 0, 0, 0, 1, ~
## $ ‘Hip Replacement Pre-Op Q Sudden Pain‘           <dbl> 0, 1, 0, 0, 0, 1, 4, ~
## $ ‘Hip Replacement Pre-Op Q Night Pain‘            <dbl> 2, 0, 1, 0, 0, 1, 1, ~
## $ ‘Hip Replacement Pre-Op Q Washing‘               <dbl> 3, 1, 1, 2, 2, 4, 4, ~
## $ ‘Hip Replacement Pre-Op Q Transport‘             <dbl> 2, 1, 1, 0, 1, 2, 2, ~
## $ ‘Hip Replacement Pre-Op Q Dressing‘              <dbl> 1, 0, 1, 0, 1, 4, 2, ~
## $ ‘Hip Replacement Pre-Op Q Shopping‘              <dbl> 3, 2, 0, 0, 0, 0, 3, ~
## $ ‘Hip Replacement Pre-Op Q Walking‘               <dbl> 2, 0, 1, 1, 1, 3, 3, ~
## $ ‘Hip Replacement Pre-Op Q Limping‘               <dbl> 2, 0, 0, 1, 0, 0, 0, ~
## $ ‘Hip Replacement Pre-Op Q Stairs‘                <dbl> 2, 1, 1, 1, 1, 2, 4, ~
## $ ‘Hip Replacement Pre-Op Q Standing‘              <dbl> 1, 1, 1, 2, 1, 1, 4, ~
## $ ‘Hip Replacement Pre-Op Q Work‘                  <dbl> 1, 1, 0, 1, 0, 0, 4, ~
## $ ‘Hip Replacement Pre-Op Q Score‘                 <dbl> 20, 8, 7, 8, 7, 18, 3~
## $ ‘Hip Replacement Post-Op Q Pain‘                 <dbl> 3, 4, 2, 2, 4, 2, 2, ~
## $ ‘Hip Replacement Post-Op Q Sudden Pain‘          <dbl> 4, 4, 4, 2, 2, 2, 4, ~
## $ ‘Hip Replacement Post-Op Q Night Pain‘           <dbl> 4, 4, 4, 1, 4, 2, 4, ~
## $ ‘Hip Replacement Post-Op Q Washing‘              <dbl> 4, 3, 3, 4, 3, 4, 4, ~
## $ ‘Hip Replacement Post-Op Q Transport‘            <dbl> 4, 4, 2, 3, 3, 2, 4, ~
## $ ‘Hip Replacement Post-Op Q Dressing‘             <dbl> 2, 4, 3, 3, 4, 4, 3, ~
## $ ‘Hip Replacement Post-Op Q Shopping‘             <dbl> 4, 2, 0, 3, 2, 0, 4, ~
## $ ‘Hip Replacement Post-Op Q Walking‘              <dbl> 4, 3, 1, 4, 3, 2, 4, ~
## $ ‘Hip Replacement Post-Op Q Limping‘              <dbl> 3, 1, 1, 4, 2, 0, 3, ~
## $ ‘Hip Replacement Post-Op Q Stairs‘               <dbl> 4, 1, 1, 3, 2, 4, 4, ~
## $ ‘Hip Replacement Post-Op Q Standing‘             <dbl> 3, 4, 3, 3, 4, 2, 4, ~
## $ ‘Hip Replacement Post-Op Q Work‘                 <dbl> 4, 4, 2, 4, 2, 2, 3, ~
## $ ‘Hip Replacement Post-Op Q Score‘                <dbl> 43, 38, 26, 36, 35, 2~
## $ ‘Hip Replacement OHS Post-Op Q Predicted‘        <dbl> 42.20017, 35.29577, 2~
```

Plot ‘EQ-5D Index’ scores pre and post operation for each gender

Select variables

I need gender, pre and post EQ-5D Index

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

Deal with 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
```

```
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
```

Make data tidy

```
head(gender_EQ5D_noNA)
```

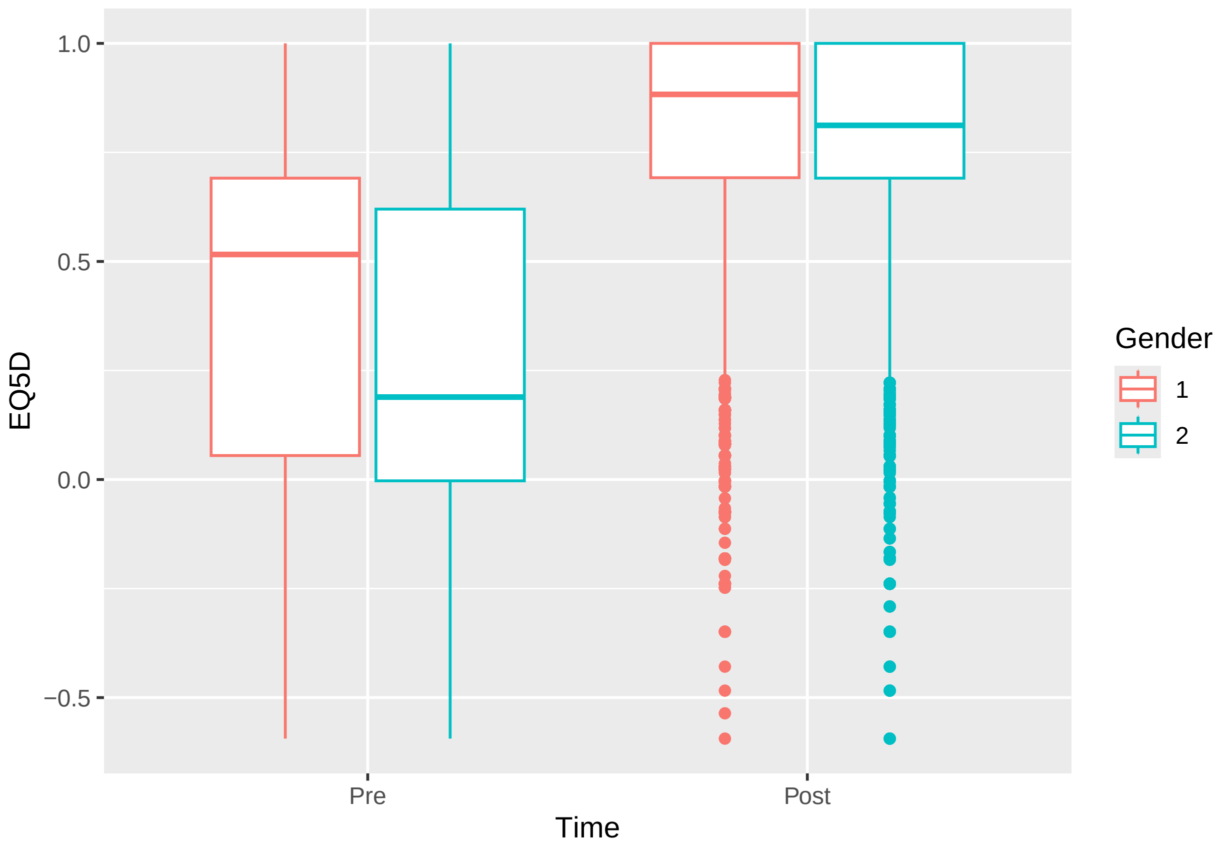
```
## # A tibble: 6 x 3
##   Gender EQ5D_Pre EQ5D_Post
##   <chr>     <dbl>     <dbl>
## 1 1        -0.016     0.516
## 2 1         0.159     0.743
## 3 1         0.03      0.727
## 4 1         0.587     0.85
## 5 1         0.623     0.796
## 6 1         0.691     1
```

```
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 orig
               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 value
               )

head(tidy_gender_EQ5D_noNA)
```

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

Answer the question

```
# 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))
```



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

Select variable

```
stroke <- hip_data %>%
  select(`Stroke`)

head(stroke)
```

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

Deal with missing data

```
stroke$Stroke %>% unique()
```

```
## [1] 9 1
```

```
#Only contains 9 or 1 and 1 means yes
stroke_noNA <- stroke %>%
  drop_na() %>%
  filter(Stroke !='9')

table(stroke_noNA$Stroke)
```

```
##
##   1
## 400
```

```
summary(stroke_noNA)
```

```
##      Stroke
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
```

Make data tidy

Stroke has only one variable, it is tidy

Answer the question

```
length(stroke_noNA$Stroke)
```

```
## [1] 400
```

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

Select variables

```
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
```

Deal with missing data

```
activity$Activity_Pre %>% unique()
```

```
## [1] 9 3 2 1
```

```
activity$Activity_Post %>% unique()
```

```
## [1] 2 9 3 1
```

```
#9 is missing
activity_noNA <- activity %>%
  drop_na() %>%
  filter(Activity_Pre != '9') %>%
  filter(Activity_Post != '9')

table(activity_noNA$Activity_Pre)
```

```
##
##     1     2     3
##  1607 20241  5386
```

```
table(activity_noNA$Activity_Post)
```

```
##
##     1     2     3
## 15932 10477   825
```

```
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
```

Make 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',   # the name of the column to create from the data stored in the orig
               names_prefix = 'Activity_',  # remove this text from the start of each variable name
               values_to = 'Activity' # the name of the column to create from the data stored in cell v
               )

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
```