

Practice week 6

Sophie Charlotte Zeiz

2024-10-29

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)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- 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

Data originates from csv file "Hip Replacement CCG 1819". It contains data from the Provisional Patient Reported Outcome Measures (PROMs) in England for Hip and Knee Replacement Procedures April 2018 to March 2019 by the NHS England.

```
hip_data <- read.csv("Hip Replacement CCG 1819.csv")
```

Prepare the data

```
glimpse(hip_data)
```

```

## Rows: 28,920
## Columns: 81
## $ Provider.Code          <chr> "00C", "00C", "00C", "0~
## $ Procedure              <chr> "Hip Replacement", "Hip~
## $ Revision.Flag          <int> 0, 0, 1, 1, 0, 0, 0, 0,~
## $ Year                    <chr> "2018/19", "2018/19", "~
## $ Age.Band                <chr> "*", "*", "*", "*", "~
## $ Gender                  <chr> "*", "*", "*", "*", "~
## $ Pre.Op.Q.Assisted       <int> 2, 2, 1, 2, 2, 2, 2, 2,~
## $ Pre.Op.Q.Assisted.By    <int> 0, 0, 0, 0, 0, 0, 0, 0,~
## $ Pre.Op.Q.Symptom.Period <int> 4, 2, 4, 1, 2, 1, 1, 2,~
## $ Pre.Op.Q.Previous.Surgery <int> 2, 1, 1, 1, 2, 2, 1, 2,~
## $ Pre.Op.Q.Living.Arrangements <int> 1, 1, 2, 2, 1, 2, 1, 2,~
## $ Pre.Op.Q.Disability     <int> 9, 1, 1, 1, 2, 1, 2, 1,~
## $ Heart.Disease           <int> 9, 9, 9, 9, 9, 9, 9, 1,~
## $ High.Bp                 <int> 9, 9, 9, 9, 9, 1, 9, 1,~
## $ Stroke                  <int> 9, 9, 9, 9, 9, 9, 1, 9,~
## $ Circulation             <int> 9, 9, 9, 9, 1, 9, 9, 9,~
## $ Lung.Disease            <int> 9, 9, 9, 9, 9, 9, 9, 9,~
## $ Diabetes                <int> 9, 9, 9, 9, 9, 9, 9, 1,~
## $ Kidney.Disease          <int> 9, 9, 9, 9, 9, 1, 9, 1,~
## $ Nervous.System          <int> 9, 9, 9, 9, 9, 9, 9, 9,~
## $ Liver.Disease           <int> 9, 9, 9, 9, 9, 9, 1, 9,~
## $ Cancer                  <int> 9, 9, 9, 9, 9, 9, 1, 9,~
## $ Depression              <int> 9, 9, 9, 1, 9, 9, 9, 9,~
## $ Arthritis               <int> 9, 1, 1, 1, 1, 1, 9, 9,~
## $ Pre.Op.Q.Mobility        <int> 2, 2, 9, 2, 2, 2, 2, 1,~
## $ Pre.Op.Q.Self.Care       <int> 1, 2, 9, 1, 2, 1, 1, 2,~
## $ Pre.Op.Q.Activity        <int> 9, 3, 9, 3, 3, 2, 2, 2,~
## $ Pre.Op.Q.Discomfort      <int> 9, 3, 9, 3, 3, 3, 2, 2,~
## $ Pre.Op.Q.Anxiety         <int> 9, 1, 9, 2, 3, 1, 1, 2,~
## $ Pre.Op.Q.EQ5D.Index.Profile <int> 21999, 22331, 99999, 21~
## $ Pre.Op.Q.EQ5D.Index     <dbl> NA, -0.003, NA, 0.030, ~
## $ Post.Op.Q.Assisted       <int> 2, 2, 1, 2, 2, 2, 1, 2,~
## $ Post.Op.Q.Assisted.By    <int> 9, 9, 1, 9, 9, 9, 1, 9,~
## $ Post.Op.Q.Living.Arrangements <int> 1, 1, 2, 2, 1, 2, 1, 9,~
## $ Post.Op.Q.Disability     <int> 2, 9, 1, 2, 1, 2, 2, 1,~
## $ Post.Op.Q.Mobility        <int> 2, 9, 2, 1, 2, 2, 1, 1,~
## $ Post.Op.Q.Self.Care       <int> 2, 1, 2, 1, 1, 1, 1, 1,~
## $ Post.Op.Q.Activity        <int> 2, 9, 3, 1, 2, 2, 1, 1,~
## $ Post.Op.Q.Discomfort      <int> 2, 1, 3, 2, 2, 2, 1, 2,~
## $ Post.Op.Q.Anxiety         <int> 2, 1, 2, 1, 2, 1, 1, 1,~
## $ Post.Op.Q.Satisfaction    <int> 2, 3, 2, 1, 3, 1, 1, 9,~
## $ Post.Op.Q.Sucess         <int> 1, 1, 1, 1, 2, 2, 1, 9,~
## $ Post.Op.Q.Allergy         <int> 2, 2, 2, 2, 2, 9, 9, 9,~
## $ Post.Op.Q.Bleeding        <int> 2, 2, 2, 2, 2, 9, 9, 9,~
## $ Post.Op.Q.Wound           <int> 2, 2, 1, 2, 2, 9, 9, 9,~
## $ Post.Op.Q.Urine           <int> 2, 2, 2, 2, 2, 1, 9, 9,~
## $ Post.Op.Q.Further.Surgery <int> 2, 2, 1, 2, 2, 2, 2, 9,~
## $ Post.Op.Q.Readmitted      <int> 2, 2, 1, 2, 2, 2, 2, 9,~
## $ Post.Op.Q.EQ5D.Index.Profile <int> 22222, 91911, 22332, 11~
## $ Post.Op.Q.EQ5D.Index     <dbl> 0.516, NA, -0.074, 0.79~
## $ Hip.Replacement.EQ5D.Index.Post.Op.Q.Predicted <dbl> NA, NA, NA, 0.5154424, ~
## $ Pre.Op.Q.EQ.VAS          <int> 999, 999, 999, 50, 30, ~

```

```
## $ Post.Op.Q.EQ.VAS <int> 70, 999, 80, 90, 70, 60~
## $ Hip.Replacement.EQ.VAS.Post.Op.Q.Predicted <dbl> NA, NA, NA, 60.05266, 7~
## $ Hip.Replacement.Pre.Op.Q.Pain <int> 1, 0, 0, 0, 0, 0, 1, 2,~
## $ Hip.Replacement.Pre.Op.Q.Sudden.Pain <int> 0, 1, 0, 0, 0, 1, 4, 3,~
## $ Hip.Replacement.Pre.Op.Q.Night.Pain <int> 2, 0, 1, 0, 0, 1, 1, 4,~
## $ Hip.Replacement.Pre.Op.Q.Washing <int> 3, 1, 1, 2, 2, 4, 4, 0,~
## $ Hip.Replacement.Pre.Op.Q.Transport <int> 2, 1, 1, 0, 1, 2, 2, 3,~
## $ Hip.Replacement.Pre.Op.Q.Dressing <int> 1, 0, 1, 0, 1, 4, 2, 0,~
## $ Hip.Replacement.Pre.Op.Q.Shopping <int> 3, 2, 0, 0, 0, 0, 3, 0,~
## $ Hip.Replacement.Pre.Op.Q.Walking <int> 2, 0, 1, 1, 1, 3, 3, 4,~
## $ Hip.Replacement.Pre.Op.Q.Limping <int> 2, 0, 0, 1, 0, 0, 0, 3,~
## $ Hip.Replacement.Pre.Op.Q.Stairs <int> 2, 1, 1, 1, 1, 2, 4, 3,~
## $ Hip.Replacement.Pre.Op.Q.Standing <int> 1, 1, 1, 2, 1, 1, 4, 4,~
## $ Hip.Replacement.Pre.Op.Q.Work <int> 1, 1, 0, 1, 0, 0, 4, 2,~
## $ Hip.Replacement.Pre.Op.Q.Score <int> 20, 8, 7, 8, 7, 18, 32,~
## $ Hip.Replacement.Post.Op.Q.Pain <int> 3, 4, 2, 2, 4, 2, 2, 9,~
## $ Hip.Replacement.Post.Op.Q.Sudden.Pain <int> 4, 4, 4, 2, 2, 2, 4, 4,~
## $ Hip.Replacement.Post.Op.Q.Night.Pain <int> 4, 4, 4, 1, 4, 2, 4, 4,~
## $ Hip.Replacement.Post.Op.Q.Washing <int> 4, 3, 3, 4, 3, 4, 4, 9,~
## $ Hip.Replacement.Post.Op.Q.Transport <int> 4, 4, 2, 3, 3, 2, 4, 3,~
## $ Hip.Replacement.Post.Op.Q.Dressing <int> 2, 4, 3, 3, 4, 4, 3, 9,~
## $ Hip.Replacement.Post.Op.Q.Shopping <int> 4, 2, 0, 3, 2, 0, 4, 0,~
## $ Hip.Replacement.Post.Op.Q.Walking <int> 4, 3, 1, 4, 3, 2, 4, 4,~
## $ Hip.Replacement.Post.Op.Q.Limping <int> 3, 1, 1, 4, 2, 0, 3, 4,~
## $ Hip.Replacement.Post.Op.Q.Stairs <int> 4, 1, 1, 3, 2, 4, 4, 4,~
## $ Hip.Replacement.Post.Op.Q.Standing <int> 3, 4, 3, 3, 4, 2, 4, 4,~
## $ Hip.Replacement.Post.Op.Q.Work <int> 4, 4, 2, 4, 2, 2, 3, 4,~
## $ Hip.Replacement.Post.Op.Q.Score <int> 43, 38, 26, 36, 35, 26,~
## $ Hip.Replacement.OHS.Post.Op.Q.Predicted <dbl> 42.20017, 35.29577, 23.~
```

Select EQ-5D scores pre and post and gender

```
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)
```

```
##   Gender EQ5D_Pre EQ5D_Post
## 1    *      NA      0.516
## 2    *   -0.003      NA
## 3    *      NA    -0.074
## 4    *    0.030    0.796
## 5    *   -0.239    0.620
## 6    *    0.159    0.691
```

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
```

```
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

```
head(gender_eq5d_noNa)
```

```
##   Gender EQ5D_Pre EQ5D_Post
## 1      1   -0.016    0.516
## 2      1    0.159    0.743
## 3      1    0.030    0.727
## 4      1    0.587    0.850
## 5      1    0.623    0.796
## 6      1    0.691    1.000
```

```
tidy_gender_eq5d_noNa <- gender_eq5d_noNa %>%
  pivot_longer(c(EQ5D_Pre, EQ5D_Post),
    names_to = 'Time',
    names_prefix = 'EQ5D_',
    values_to = 'EQ5D'
  )

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
```

Rename Variable for Gender

```
tidy_gender_eq5d_noNa$Gender <- factor(tidy_gender_eq5d_noNa$Gender, labels=c("male", "female"))

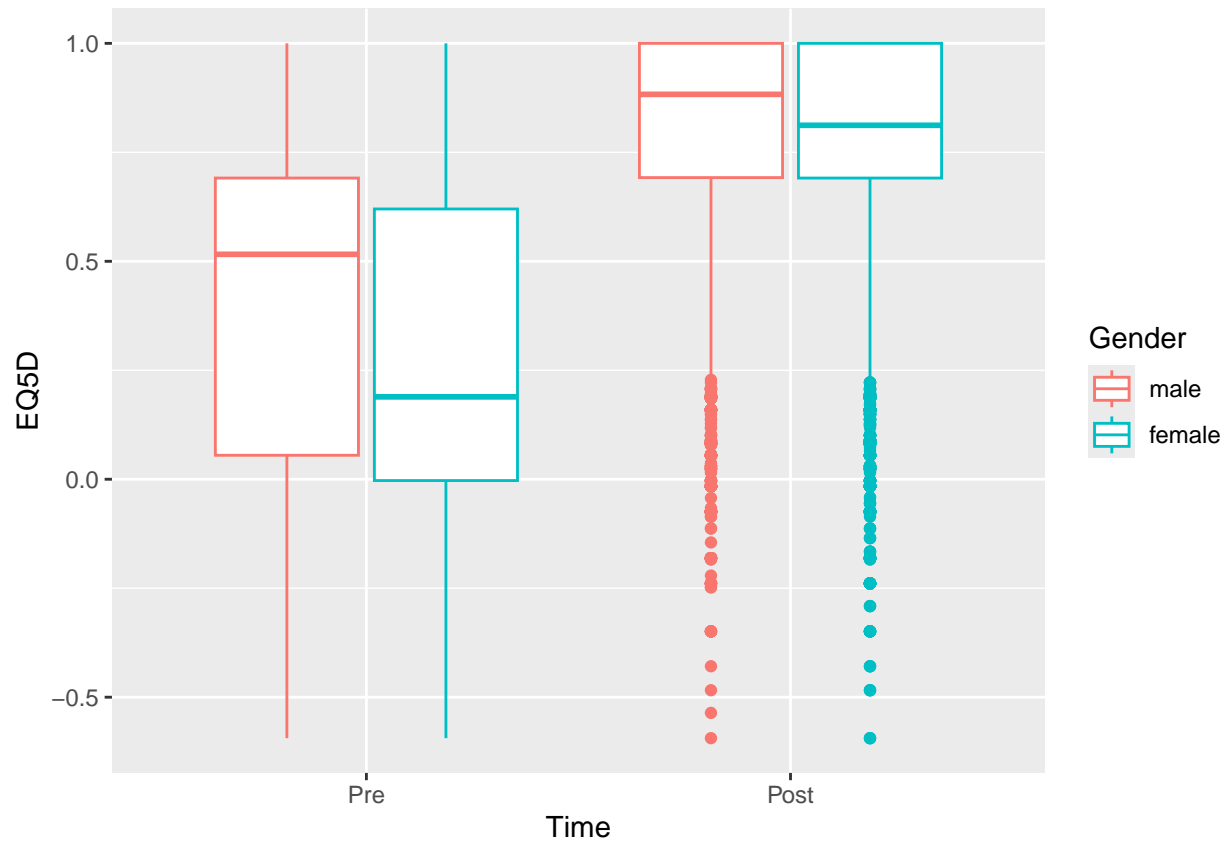
head(tidy_gender_eq5d_noNa)
```

```
## # A tibble: 6 x 3
##   Gender Time    EQ5D
##   <fct> <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 EQ5D Index pre and post operation for each gender

```
tidy_gender_eq5d_noNa$Time <- factor(tidy_gender_eq5d_noNa$Time, levels = c('Pre', 'Post'))

tidy_gender_eq5d_noNa %>%
  ggplot()+
  geom_boxplot(aes(x = Time, y = EQ5D, colour = Gender))
```



```
##Stroke
```

```
stroke_data <- hip_data %>%
  select(Stroke) %>%
  group_by(Stroke) %>%
  count(Stroke)

stroke_data$Stroke <- factor(stroke_data$Stroke, labels=c("yes","missing"))

head(stroke_data)
```

```
## # A tibble: 2 x 2
## # Groups:   Stroke [2]
##   Stroke      n
##   <fct>   <int>
## 1 yes       400
## 2 missing 28520
```

400 people were told their symptoms have been caused from a stroke

Pre and post operation activity levels *clean and tidy*

```
activity_data <- 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_data)

```

```

##   Activity_Pre Activity_Post
## 1           9           2
## 2           3           9
## 3           9           3
## 4           3           1
## 5           3           2
## 6           2           2

```

```
summary(activity_data)
```

```

##   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

```

```

tidy_activity_data <- activity_data %>%
  pivot_longer(c(Activity_Pre, Activity_Post),
               names_to = 'Time',
               names_prefix = 'Activity_',
               values_to = 'Activity'
  )
table(tidy_activity_data)

```

```

##           Activity
## Time      1      2      3      9
##  Post 16602 10918   858   542
##  Pre   1625 20622  5499  1174

```

Order table

```

tidy_activity_data$Time <- factor(tidy_activity_data$Time, levels = c('Pre', 'Post'))
head(tidy_activity_data)

```

```

## # A tibble: 6 x 2
##   Time Activity
##   <fct>   <int>
## 1 Pre       9
## 2 Post       2
## 3 Pre       3
## 4 Post       9
## 5 Pre       9
## 6 Post       3

```