PU5558-PU5567-computer-lab-report

Anonymous

Load necessary packages

We begin by loading the nessicary packages required to carry out all comands needed in this report. Tidyverse, as we learnt in our previous module is required for our general data science, along with "ggplots" for visualisation, and corrplot for visualising correlation matrices and tinymodels which was introduced to us this term is required for machine learning - our main focus this term.

```
#Install 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     1.0.4
## -- 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(tidymodels)
```

```
## -- Attaching packages --------------------------------------- tidymodels 1.3.0 --
## v broom        1.0.7     v rsample      1.3.0
## v dials        1.4.0     v tune         1.3.0
## v infer        1.0.7     v workflows    1.2.0
## v modeldata    1.4.0     v workflowsets 1.1.0
## v parsnip      1.3.1     v yardstick    1.3.2
## v recipes      1.2.1
## -- Conflicts ------------------------------------------ tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
```

```
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
library(ranger)
library(devtools)
```

```
## Loading required package: usethis
##
## Attaching package: 'devtools'
##
## The following object is masked from 'package:recipes':
##
##     check
```

Load chosen dataset

We were provided two seperate data sets containing "Provisional Patient Record Outcomes" to select from to answer our question for this assignment. Out of the two I have selected "Knee replacement CCG 2021" which can be found at the following address: https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/hip-and-knee-replacement-procedures-april-2020-to-march-2021

To load this data set into my studio to then work with I will be using the "read_csv" function.

```
#Read in data
Knee_Data <-read_csv ("~/Downloads/Knee Replacement CCG 2021.csv")
```

```
## Rows: 5422 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.
```

```
glimpse(Knee_Data)
```

```
## Rows: 5,422
## Columns: 81
## $ `Provider Code`                                <chr> "00C", "00C", "00C"~
## $ Procedure                                      <chr> "Knee Replacement",~
## $ `Revision Flag`                                <dbl> 0, 0, 0, 0, 0, 0, 0~
## $ Year                                           <chr> "2020/21", "2020/21~
## $ `Age Band`                                     <chr> "*", "*", "*", "*",~
## $ Gender                                         <chr> "*", "*", "*", "*",~
## $ `Pre-Op Q Assisted`                            <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Pre-Op Q Assisted By`                         <dbl> 0, 0, 0, 0, 0, 0, 0~
## $ `Pre-Op Q Symptom Period`                      <dbl> 2, 4, 4, 2, 4, 2, 2~
## $ `Pre-Op Q Previous Surgery`                    <dbl> 2, 1, 2, 2, 2, 2, 2~
## $ `Pre-Op Q Living Arrangements`                 <dbl> 2, 2, 1, 1, 1, 1, 1~
## $ `Pre-Op Q Disability`                          <dbl> 1, 2, 2, 2, 1, 2, 2~
## $ `Heart Disease`                                <dbl> 9, 9, 9, 9, 9, 1, 9~
## $ `High Bp`                                      <dbl> 1, 9, 9, 1, 1, 1, 9~
## $ Stroke                                         <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ Circulation                                    <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ `Lung Disease`                                 <dbl> 9, 9, 9, 1, 9, 9, 9~
## $ Diabetes                                       <dbl> 9, 9, 9, 1, 9, 9, 9~
## $ `Kidney Disease`                               <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ `Nervous System`                               <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ `Liver Disease`                                <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ Cancer                                         <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ Depression                                     <dbl> 9, 9, 9, 9, 9, 9, 9~
```

```
## $ Arthritis                                          <dbl> 9, 1, 1, 1, 1, 1, 1~
## $ `Pre-Op Q Mobility`                                <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Pre-Op Q Self-Care`                               <dbl> 1, 1, 1, 1, 1, 1, 1~
## $ `Pre-Op Q Activity`                                <dbl> 3, 2, 3, 2, 1, 2, 2~
## $ `Pre-Op Q Discomfort`                              <dbl> 2, 2, 3, 2, 1, 2, 2~
## $ `Pre-Op Q Anxiety`                                 <dbl> 1, 1, 2, 2, 1, 2, 1~
## $ `Pre-Op Q EQ5D Index Profile`                      <dbl> 21321, 21221, 21332~
## $ `Pre-Op Q EQ5D Index`                              <dbl> 0.364, 0.691, 0.030~
## $ `Post-Op Q Assisted`                               <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Post-Op Q Assisted By`                            <dbl> 9, 9, 9, 9, 9, 9, 9~
## $ `Post-Op Q Living Arrangements`                    <dbl> 2, 2, 1, 1, 1, 1, 1~
## $ `Post-Op Q Disability`                             <dbl> 1, 2, 2, 2, 2, 1, 2~
## $ `Post-Op Q Mobility`                               <dbl> 1, 1, 1, 1, 1, 2, 1~
## $ `Post-Op Q Self-Care`                              <dbl> 2, 1, 1, 1, 1, 2, 1~
## $ `Post-Op Q Activity`                               <dbl> 1, 1, 1, 1, 1, 2, 1~
## $ `Post-Op Q Discomfort`                             <dbl> 1, 1, 1, 1, 1, 2, 2~
## $ `Post-Op Q Anxiety`                                <dbl> 2, 1, 1, 1, 1, 1, 1~
## $ `Post-Op Q Satisfaction`                           <dbl> 2, 3, 1, 2, 1, 4, 3~
## $ `Post-Op Q Sucess`                                 <dbl> 1, 1, 1, 1, 1, 4, 1~
## $ `Post-Op Q Allergy`                                <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Post-Op Q Bleeding`                               <dbl> 2, 2, 2, 2, 1, 2, 2~
## $ `Post-Op Q Wound`                                  <dbl> 1, 2, 2, 2, 1, 2, 2~
## $ `Post-Op Q Urine`                                  <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Post-Op Q Further Surgery`                        <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Post-Op Q Readmitted`                             <dbl> 2, 2, 2, 2, 2, 2, 2~
## $ `Post-Op Q EQ5D Index Profile`                     <dbl> 12112, 11111, 11111~
## $ `Post-Op Q EQ5D Index`                             <dbl> 0.744, 1.000, 1.000~
## $ `Knee Replacement EQ 5D Index Post-Op Q Predicted` <dbl> 0.6621424, 0.740953~
## $ `Pre-Op Q EQ VAS`                                  <dbl> 85, 50, 46, 60, 60,~
## $ `Post-Op Q EQ VAS`                                 <dbl> 60, 100, 90, 95, 90~
## $ `Knee Replacement EQ VAS_Post-Op Q Predicted`      <dbl> 75.63073, 66.38606,~
## $ `Knee Replacement Pre-Op Q Pain`                   <dbl> 1, 1, 0, 0, 1, 1, 1~
## $ `Knee Replacement Pre-Op Q Night Pain`             <dbl> 1, 2, 2, 2, 2, 3, 0~
## $ `Knee Replacement Pre-Op Q Washing`                <dbl> 3, 4, 3, 4, 3, 3, 3~
## $ `Knee Replacement Pre-Op Q Transport`              <dbl> 1, 4, 2, 4, 2, 2, 2~
## $ `Knee Replacement Pre-Op Q Walking`                <dbl> 3, 3, 2, 3, 3, 1, 2~
## $ `Knee Replacement Pre-Op Q Standing`               <dbl> 3, 2, 2, 3, 3, 2, 2~
## $ `Knee Replacement Pre-Op Q Limping`                <dbl> 0, 3, 0, 0, 3, 1, 1~
## $ `Knee Replacement Pre-Op Q Kneeling`               <dbl> 0, 0, 0, 2, 1, 1, 2~
## $ `Knee Replacement Pre-Op Q Work`                   <dbl> 2, 2, 1, 3, 2, 1, 2~
## $ `Knee Replacement Pre-Op Q Confidence`             <dbl> 3, 3, 3, 2, 1, 1, 2~
## $ `Knee Replacement Pre-Op Q Shopping`               <dbl> 2, 4, 0, 2, 3, 0, 2~
## $ `Knee Replacement Pre-Op Q Stairs`                 <dbl> 1, 4, 2, 2, 3, 2, 2~
## $ `Knee Replacement Pre-Op Q Score`                  <dbl> 20, 32, 17, 27, 27,~
## $ `Knee Replacement Post-Op Q Pain`                  <dbl> 4, 3, 3, 4, 3, 1, 2~
## $ `Knee Replacement Post-Op Q Night Pain`            <dbl> 4, 4, 3, 4, 4, 3, 2~
## $ `Knee Replacement Post-Op Q Washing`               <dbl> 3, 4, 4, 4, 4, 1, 4~
## $ `Knee Replacement Post-Op Q Transport`             <dbl> 2, 4, 4, 4, 4, 3, 4~
## $ `Knee Replacement Post-Op Q Walking`               <dbl> 4, 4, 4, 4, 4, 1, 4~
## $ `Knee Replacement Post-Op Q Standing`              <dbl> 3, 4, 4, 4, 4, 1, 3~
## $ `Knee Replacement Post-Op Q Limping`               <dbl> 3, 4, 4, 4, 4, 1, 4~
## $ `Knee Replacement Post-Op Q Kneeling`              <dbl> 0, 4, 0, 3, 2, 0, 2~
## $ `Knee Replacement Post-Op Q Work`                  <dbl> 3, 3, 4, 4, 4, 1, 4~
## $ `Knee Replacement Post-Op Q Confidence`            <dbl> 4, 4, 4, 4, 4, 1, 4~
```

```
## $ `Knee Replacement Post-Op Q Shopping`              <dbl> 4, 4, 4, 4, 4, 0, 4~
## $ `Knee Replacement Post-Op Q Stairs`                <dbl> 3, 4, 4, 4, 3, 1, 4~
## $ `Knee Replacement Post-Op Q Score`                 <dbl> 37, 46, 42, 47, 44,~
## $ `Knee Replacement OKS Post-Op Q Predicted`         <dbl> 34.02619, 36.88715,~
```

Dataset description

Suitable machine learning algorithm for three questions:

#1. Before the operation, can we estimate the post-operative EQ5D index for a patient?

Answer: When looking to estimate post-operative EQ5D index we are looking at a regression problem as the EQ5D is a continuos value outcome, therefore require supervised machine learning to solve the issue. This task requires to use a dataset of patient reported outcomes and other relevant information to train a model to predict the EQ5D score. Therefore I would choose to utilise a random forest regressor as it has the ability to handle both catagoricol and numerical predictors and is good when handling data which may contain missing values and outliers.

When looking into which model would be best suited for the task I also investigated the use of linear regression however for this task in particular it is important the model does not assume a linear relationship between variables and allows for more flexibility when working with more complex interactions in patient data, e.g age or pre-existing conditions.

Overall, I beleive Random forest has a better balance of predictive performance with practical usability for this task making it my choice for estimating post-operative EQ5D- index.

#2. Before the operation, can we predict how much pain a patient will have after the operation?

Answers: In the 'Knee replacement data set' we have the data needed to carry out this task in the column 'Knee Replacement Post-Op Q Pain' where the intensity of pain has been given a numerical value therefore making this a regression problem. Classification, a type of supervised machine learning method should be used in this task. Again linear regression could be used in this task however it would assume the linear relationship between predictors and pain score. I would be more inclined to use either Random forest regressor or Support vector regression in order to predict pain.

#3. Before the operation, can we calculate how many patients have had previous surgery?

Answer: Yes, we can calculate the number of patients that had prior surgeries in our data set. Our data set contains the column "Pre-Op Q Previous Surgery" in which patients answers are coded as follows; 1 = 'Yes', 2 = 'No', and 9 = 'Missing value'. Therefore we can filter for specific '1' value to calculate the number of patients who had previous surgies. In our data set this means that there were 537 patients in our data set who had previous surgery.

table(Knee_Data$`Pre-Op Q Previous Surgery`)

```
# Set seed for reproducibility
set.seed(123)

# Split the data (80% train, 20% test) stratified by pain score
Knee_split <- initial_split(Knee_Data,
                            prop = 0.8,
                            strata = `Knee Replacement Post-Op Q Pain`)

# Create training and testing datasets
knee_train <- training(Knee_split)
knee_test <- testing(Knee_split)

# Clean and prep training and testing data
knee_train_clean <- knee_train %>%
```

```r
  drop_na() %>%
  select(-any_of(c("Procedure", "Year"))) %>%                    # Drop Procedure & Year together
  mutate(across(where(is.character), as.factor))


# heck for any missing values just in case
colSums(is.na(knee_train_clean))
```

```
##                                    Provider Code
##                                                0
##                                   Revision Flag
##                                                0
##                                        Age Band
##                                                0
##                                          Gender
##                                                0
##                                 Pre-Op Q Assisted
##                                                0
##                              Pre-Op Q Assisted By
##                                                0
##                            Pre-Op Q Symptom Period
##                                                0
##                           Pre-Op Q Previous Surgery
##                                                0
##                         Pre-Op Q Living Arrangements
##                                                0
##                                  Pre-Op Q Disability
##                                                0
##                                     Heart Disease
##                                                0
##                                           High Bp
##                                                0
##                                           Stroke
##                                                0
##                                      Circulation
##                                                0
##                                     Lung Disease
##                                                0
##                                         Diabetes
##                                                0
##                                    Kidney Disease
##                                                0
##                                   Nervous System
##                                                0
##                                    Liver Disease
##                                                0
##                                           Cancer
##                                                0
##                                       Depression
##                                                0
##                                        Arthritis
##                                                0
##                                 Pre-Op Q Mobility
##                                                0
```

```
##                                Pre-Op Q Self-Care
##                                                 0
##                                 Pre-Op Q Activity
##                                                 0
##                               Pre-Op Q Discomfort
##                                                 0
##                                  Pre-Op Q Anxiety
##                                                 0
##                        Pre-Op Q EQ5D Index Profile
##                                                 0
##                                Pre-Op Q EQ5D Index
##                                                 0
##                                Post-Op Q Assisted
##                                                 0
##                              Post-Op Q Assisted By
##                                                 0
##                       Post-Op Q Living Arrangements
##                                                 0
##                                Post-Op Q Disability
##                                                 0
##                                 Post-Op Q Mobility
##                                                 0
##                                Post-Op Q Self-Care
##                                                 0
##                                 Post-Op Q Activity
##                                                 0
##                               Post-Op Q Discomfort
##                                                 0
##                                  Post-Op Q Anxiety
##                                                 0
##                             Post-Op Q Satisfaction
##                                                 0
##                                  Post-Op Q Sucess
##                                                 0
##                                 Post-Op Q Allergy
##                                                 0
##                                Post-Op Q Bleeding
##                                                 0
##                                   Post-Op Q Wound
##                                                 0
##                                   Post-Op Q Urine
##                                                 0
##                          Post-Op Q Further Surgery
##                                                 0
##                                Post-Op Q Readmitted
##                                                 0
##                        Post-Op Q EQ5D Index Profile
##                                                 0
##                                Post-Op Q EQ5D Index
##                                                 0
## Knee Replacement EQ 5D Index Post-Op Q Predicted
##                                                 0
##                                     Pre-Op Q EQ VAS
##                                                 0
```

```
##                                Post-Op Q EQ VAS
##                                               0
##      Knee Replacement EQ VAS_Post-Op Q Predicted
##                                               0
##                  Knee Replacement Pre-Op Q Pain
##                                               0
##              Knee Replacement Pre-Op Q Night Pain
##                                               0
##                Knee Replacement Pre-Op Q Washing
##                                               0
##               Knee Replacement Pre-Op Q Transport
##                                               0
##                Knee Replacement Pre-Op Q Walking
##                                               0
##               Knee Replacement Pre-Op Q Standing
##                                               0
##                Knee Replacement Pre-Op Q Limping
##                                               0
##               Knee Replacement Pre-Op Q Kneeling
##                                               0
##                   Knee Replacement Pre-Op Q Work
##                                               0
##              Knee Replacement Pre-Op Q Confidence
##                                               0
##                Knee Replacement Pre-Op Q Shopping
##                                               0
##                 Knee Replacement Pre-Op Q Stairs
##                                               0
##                  Knee Replacement Pre-Op Q Score
##                                               0
##                  Knee Replacement Post-Op Q Pain
##                                               0
##             Knee Replacement Post-Op Q Night Pain
##                                               0
##                Knee Replacement Post-Op Q Washing
##                                               0
##               Knee Replacement Post-Op Q Transport
##                                               0
##                Knee Replacement Post-Op Q Walking
##                                               0
##               Knee Replacement Post-Op Q Standing
##                                               0
##                Knee Replacement Post-Op Q Limping
##                                               0
##               Knee Replacement Post-Op Q Kneeling
##                                               0
##                   Knee Replacement Post-Op Q Work
##                                               0
##             Knee Replacement Post-Op Q Confidence
##                                               0
##               Knee Replacement Post-Op Q Shopping
##                                               0
##                 Knee Replacement Post-Op Q Stairs
##                                               0
```

```
##                 Knee Replacement Post-Op Q Score
##                                                0
##         Knee Replacement OKS Post-Op Q Predicted
##                                                0
```

```
# Build recipe
pain_recipe <- recipe(`Knee Replacement Post-Op Q Pain` ~ ., data = knee_train_clean) %>%
  step_dummy(all_nominal_predictors(), one_hot = TRUE) %>%
  step_zv(all_predictors())  # Remove predictors with zero variance

# specify a random forest model
rf_model <- rand_forest(mode = "regression", trees = 500) %>%
  set_engine("ranger")

# Create a workflow to combine recipe and model
rf_workflow <- workflow() %>%
  add_recipe(pain_recipe) %>%
  add_model(rf_model)

# rain the model
rf_fit <- rf_workflow %>%
  fit(data = knee_train_clean)
```

2. Selection and preprocessing of predictors

```
knee_train_clean <- knee_train %>%
  drop_na() %>%
  select(-Procedure) %>%
  mutate(across(where(is.character), as.factor))

knee_test_clean <- knee_test %>%
  drop_na() %>%
  select(-Procedure) %>%
  mutate(across(where(is.character), as.factor))

# Check for missing values just in case
colSums(is.na(knee_train_clean))
```

```
##                                    Provider Code
##                                                0
##                                    Revision Flag
##                                                0
##                                             Year
##                                                0
##                                         Age Band
##                                                0
##                                           Gender
##                                                0
##                                 Pre-Op Q Assisted
##                                                0
##                              Pre-Op Q Assisted By
##                                                0
##                            Pre-Op Q Symptom Period
##                                                0
##                          Pre-Op Q Previous Surgery
##                                                0
```

```
##                     Pre-Op Q Living Arrangements
##                                                0
##                              Pre-Op Q Disability
##                                                0
##                                    Heart Disease
##                                                0
##                                         High Bp
##                                                0
##                                          Stroke
##                                                0
##                                     Circulation
##                                                0
##                                     Lung Disease
##                                                0
##                                        Diabetes
##                                                0
##                                   Kidney Disease
##                                                0
##                                   Nervous System
##                                                0
##                                    Liver Disease
##                                                0
##                                          Cancer
##                                                0
##                                     Depression
##                                                0
##                                      Arthritis
##                                                0
##                                 Pre-Op Q Mobility
##                                                0
##                                Pre-Op Q Self-Care
##                                                0
##                                 Pre-Op Q Activity
##                                                0
##                                Pre-Op Q Discomfort
##                                                0
##                                   Pre-Op Q Anxiety
##                                                0
##                  Pre-Op Q EQ5D Index Profile
##                                                0
##                                Pre-Op Q EQ5D Index
##                                                0
##                                 Post-Op Q Assisted
##                                                0
##                               Post-Op Q Assisted By
##                                                0
##                     Post-Op Q Living Arrangements
##                                                0
##                                Post-Op Q Disability
##                                                0
##                                 Post-Op Q Mobility
##                                                0
##                                Post-Op Q Self-Care
##                                                0
```

```
##                              Post-Op Q Activity
##                                               0
##                            Post-Op Q Discomfort
##                                               0
##                               Post-Op Q Anxiety
##                                               0
##                           Post-Op Q Satisfaction
##                                               0
##                                Post-Op Q Sucess
##                                               0
##                               Post-Op Q Allergy
##                                               0
##                              Post-Op Q Bleeding
##                                               0
##                                 Post-Op Q Wound
##                                               0
##                                 Post-Op Q Urine
##                                               0
##                         Post-Op Q Further Surgery
##                                               0
##                              Post-Op Q Readmitted
##                                               0
##                      Post-Op Q EQ5D Index Profile
##                                               0
##                              Post-Op Q EQ5D Index
##                                               0
## Knee Replacement EQ 5D Index Post-Op Q Predicted
##                                               0
##                                  Pre-Op Q EQ VAS
##                                               0
##                                 Post-Op Q EQ VAS
##                                               0
##      Knee Replacement EQ VAS_Post-Op Q Predicted
##                                               0
##                    Knee Replacement Pre-Op Q Pain
##                                               0
##              Knee Replacement Pre-Op Q Night Pain
##                                               0
##                 Knee Replacement Pre-Op Q Washing
##                                               0
##               Knee Replacement Pre-Op Q Transport
##                                               0
##                 Knee Replacement Pre-Op Q Walking
##                                               0
##                Knee Replacement Pre-Op Q Standing
##                                               0
##                 Knee Replacement Pre-Op Q Limping
##                                               0
##                Knee Replacement Pre-Op Q Kneeling
##                                               0
##                    Knee Replacement Pre-Op Q Work
##                                               0
##              Knee Replacement Pre-Op Q Confidence
##                                               0
```

```
##               Knee Replacement Pre-Op Q Shopping
##                                                0
##                 Knee Replacement Pre-Op Q Stairs
##                                                0
##                  Knee Replacement Pre-Op Q Score
##                                                0
##                  Knee Replacement Post-Op Q Pain
##                                                0
##            Knee Replacement Post-Op Q Night Pain
##                                                0
##               Knee Replacement Post-Op Q Washing
##                                                0
##             Knee Replacement Post-Op Q Transport
##                                                0
##               Knee Replacement Post-Op Q Walking
##                                                0
##              Knee Replacement Post-Op Q Standing
##                                                0
##               Knee Replacement Post-Op Q Limping
##                                                0
##              Knee Replacement Post-Op Q Kneeling
##                                                0
##                  Knee Replacement Post-Op Q Work
##                                                0
##            Knee Replacement Post-Op Q Confidence
##                                                0
##              Knee Replacement Post-Op Q Shopping
##                                                0
##                Knee Replacement Post-Op Q Stairs
##                                                0
##                 Knee Replacement Post-Op Q Score
##                                                0
##         Knee Replacement OKS Post-Op Q Predicted
##                                                0
```

```r
# Build recipe (now safe to use step_dummy)
pain_recipe <- recipe(`Knee Replacement Post-Op Q Pain` ~ ., data = knee_train_clean) %>%
  step_dummy(all_nominal_predictors(), one_hot = TRUE) %>%  # Creates dummy vars
  step_zv(all_predictors())
```

3. Model specification and training

```r
# Specify a random forest model
rf_model <- rand_forest(mode = "regression", trees = 500) %>%
  set_engine("ranger")

# Create a workflow to combine recipe and model
rf_workflow <- workflow() %>%
  add_recipe(pain_recipe) %>%
  add_model(rf_model)

pain_recipe <- recipe(`Knee Replacement Post-Op Q Pain` ~ ., data = knee_train_clean) %>%
  step_dummy(all_nominal_predictors()) %>%
  step_zv(all_predictors())  # Remove zero variance predictors
```

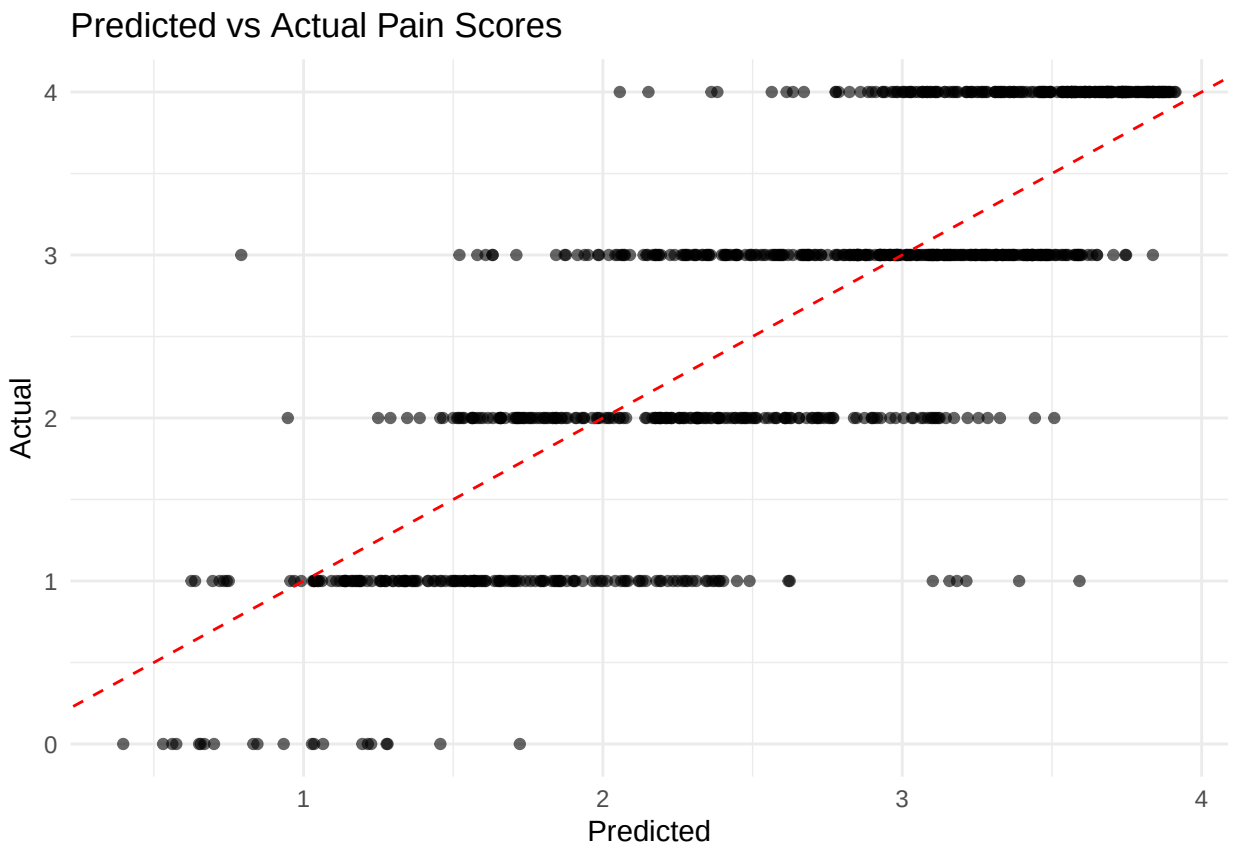
```r
# Train the random forest model on the cleaned training data
pain_recipe <- recipe(`Knee Replacement Post-Op Q Pain` ~ ., data = knee_train_clean) %>%
  step_dummy(all_nominal_predictors(), one_hot = TRUE) %>%
  step_zv(all_predictors)
```

4. Model evaluation

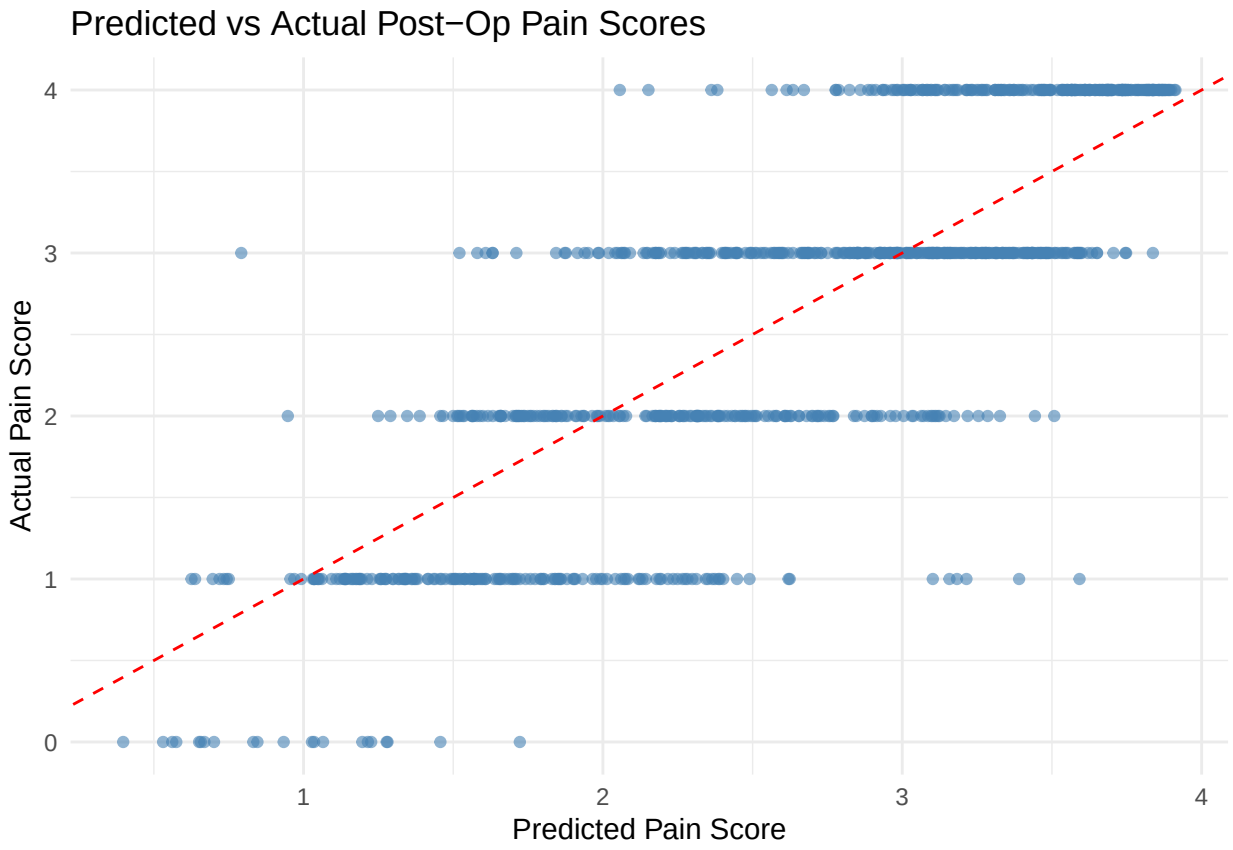
```r
# Clean the test set (just like training)
knee_test_clean <- knee_test %>%
  drop_na() %>%
  select(-any_of(c("Procedure", "Year"))) %>%
  mutate(across(where(is.character), as.factor))

# Predict on test data
rf_predictions <- predict(rf_fit, new_data = knee_test_clean) %>%
  bind_cols(knee_test_clean %>% select(`Knee Replacement Post-Op Q Pain`))

# Evaluate performance
rf_metrics <- rf_predictions %>%
  metrics(truth = `Knee Replacement Post-Op Q Pain`, estimate = .pred)

rf_metrics
```

```
## # A tibble: 3 x 3
##   .metric .estimator .estimate
##   <chr>   <chr>          <dbl>
## 1 rmse    standard       0.648
## 2 rsq     standard       0.667
## 3 mae     standard       0.516
```

```r
#Visualisation

ggplot(rf_predictions, aes(x = .pred, y = `Knee Replacement Post-Op Q Pain`)) +
  geom_point(alpha = 0.6) +
  geom_abline(color = "red", linetype = "dashed") +
  labs(title = "Predicted vs Actual Pain Scores",
       x = "Predicted",
       y = "Actual") +
  theme_minimal()
```




```
ggplot(rf_predictions, aes(x = .pred, y = `Knee Replacement Post-Op Q Pain`)) +
  geom_point(alpha = 0.6, color = "steelblue") +
  geom_abline(color = "red", linetype = "dashed") +
  labs(title = "Predicted vs Actual Post-Op Pain Scores",
       x = "Predicted Pain Score",
       y = "Actual Pain Score") +
  theme_minimal()
```



Limitations of machine learning model