First Draft Report: Group 6 Ametop Study with Dr. Louis Scheepers

21 February 2024

- Fabiola Grace

- Maggie Ruan

- Runhe Guo

- Yimin You

Summary

This study compares the efficacy of two approaches: Ametop alone and Ametop combined with a vapocoolant spray, in alleviating pain during IV insertion. In the experiment, pain levels were assessed alongside secondary outcomes. Statistical analyses revealed that the addition of vapocoolant spray may not significantly reduce discomfort in pediatric patients. However, age, number of IV attempts, and practitioner's role emerged as significant factors influencing pain levels, highlighting the importance of personalized care in pediatric anesthesia practices. These findings contribute to optimizing pain management strategies and improving patient outcomes in pediatric anesthesia.

Introduction

Pediatric patients commonly experience anxiety and discomfort during IV insertion. In this study, we aim to compare the efficacy of two pain management strategies to improve pediatric anesthesia practices and alleviate the pain associated with this procedure: Ametop alone and its combination with a vapocoolant spray. The research employed a randomized controlled trial to assess pain levels experienced by children, while also exploring secondary outcomes such as the impact of provider experience, gender, and age-related differences in pain perception. Our statistical analysis will utilize a t-test to assess the primary outcome of pain reduction due to the addition of vapocoolant spray, a linear regression analysis to investigate the impact of other factors on pain levels and pie charts to visualize the distribution of side effects.

Data description

The study retrospectively collected data from 240 pediatric patients undergoing intravenous (IV) insertion. There were 9 controlled variables to ensure the consistency of the participants in the experiment. Some participants withdrew from the study, citing reasons such as conversion to inhalation induction, Ametop-induced itchiness requiring removal. As these cases lacked sufficient information in recording the response variables, the missing variables cannot be addressed through imputations or other methods. Consequently, we excluded these 17 observations, focusing solely on the remaining 223 observations for main and secondary objective 1. The dataset has several key variables, categorized as follows:

- **Patient Demographics:** This includes variables such as age, gender, and ASA classification, which are crucial for subgroup analyses. Age (5-16 years old) is particularly important for determining if responses to pain management strategies vary among different pediatric age groups.
- **Treatment Group Randomization:** Patients are randomly assigned to one of two groups: Group 1, receiving Ametop alone (the control group), and Group 2, receiving both Ametop and the vapocoolant spray (the study group). This randomization is essential for assessing the comparative effectiveness of the two treatment approaches.
- **FPS-R Scores:** Patients rated their pain using the FPS-R scale from 0 (no pain) to 10 (worst pain possible), offering a measurable discomfort level during IV insertion. This score determines the primary outcome, comparing the effectiveness of treatment approaches.
- **Observer's Scores:** The patients' discomfort is observed and rated by a nurse on a 3-point scale: 'none,' 'slight,' or 'severe' to capture visible signs of discomfort or pain. This complements FPS-R scores, ensuring reliable pain assessment in children, as some children may be playful or may not fully understand the meaning of the pictograms provided.
- **Number of IV Attempts:** This variable records the number of attempts required to successfully insert the IV, reflecting the procedural difficulty and its potential correlation with patient discomfort.
- **Needle Gauge:** The size of the needle used, differentiated as "22" being slightly larger than "24," was accounted for in the analysis due to its potential influence on discomfort levels.
- **Practitioner's role:** The practitioner is either a resident or an anesthetist which may result in a difference in pain level felt by the patient.
- **Side Effects:** Side effects were monitored to ensure uniform treatment across participants.

Methods

Main Objective: Does the combination of Ametop with a vapocoolant spray significantly reduce the discomfort experienced by pediatric patients during IV insertions, compared to the application of Ametop alone?

The main challenge in assessing our primary objective lies in the presence of two target responses: the FPS-R scores selected by the patients and the observer's scores. Accurately measuring pain in children using FPS-R scores alone can be difficult, as some children may be playful or might not fully comprehend the pictograms provided. Thus, the observer's scores serve as an additional target response, intended to confirm the reliability of the FPS-R scores. While the patient's FPS-R score remains the focal point, its alignment with the observer's scores is necessary. Consequently, we will conduct the statistical analysis and testing in two ways: a test covering the entire dataset, and a second test limited to a subset of data where the observer's scores and the patient's FPS-R score align.

		FPS-R Score		
		0 & 2	4 & 6	8 & 10
Observer's Score	Severe	0	0	5
	Slight	24	24	12
	None	135	20	3

Figure 1: Count of matches in data

Figure 1 shows the total number of patients where the observer's scores and the patient's FPS-R score align and where it does not align. To create the subset of data where the score matches, we use the following:

- FPS-R scores 0 and 2 matched with the observer's score 'None' = 135 patients
- FPS-R scores 4 and 6 matched with the observer's score 'Slight Pain' = 24 patients
- FPS-R scores 8 and 10 matched with the observer's score 'Severe Pain' = 5 patients

Before carrying out the statistical test, we will perform exploratory data analysis to examine the distribution of the target response, the FPS-R score.
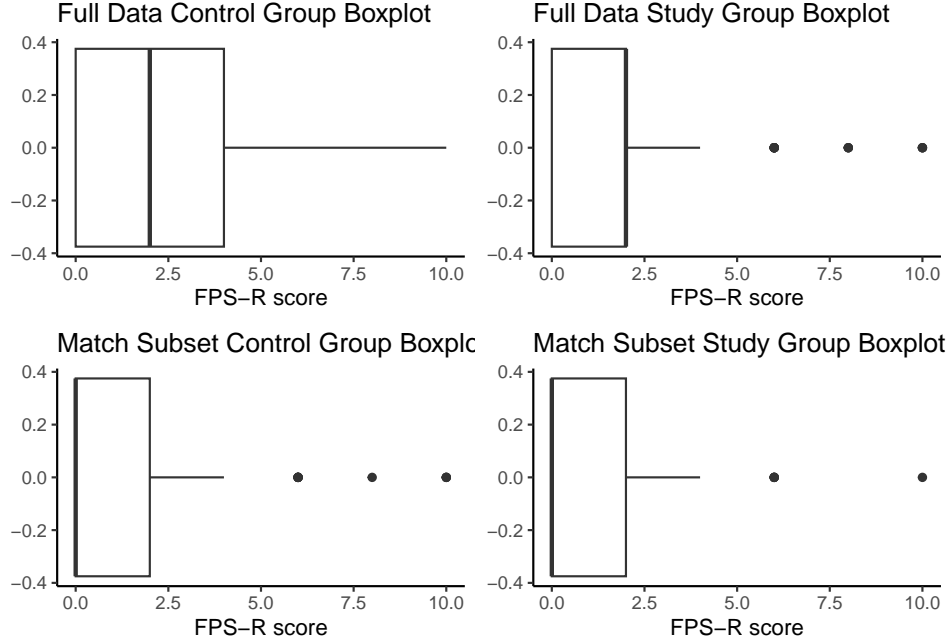
Figure 2: Boxplots of Control and Study Group for Full and Subset Data

From Figure 2, we observe that the median FPS-R score for both the control and treatment groups is 2. The treatment group's data shows a higher concentration of scores at FPS-R 0 and 2, and is right-skewed. When examining the boxplot of the subset of data where the FPS-R scores and observer's scores align, their distributions do not appear significantly different.

The statistical method used to assess whether the addition of vapocoolant spray reduces discomfort is the Welch Two Sample T-Test. This test is applied under the assumption that the differences between the control group and the study group follow a Normal distribution, as per the Central Limit Theorem (CLT). We conduct the t-test twice: first on the entire dataset, and second on the subset of data where the FPS-R scores and observer's scores match. The null hypothesis is that the mean pain level is the same for both treatment groups, while the alternative hypothesis is that the mean pain level is lower for the treatment group using the vapocoolant spray compared to the group using Ametop alone. Thus, the test is a one-sided t-test. In accordance with Dr. Louis Scheepers' recommendation, the threshold for p-value significance is set at a 5% significance level.

Secondary Objective 1: Do age, gender, the practitioner's role, the number of IV attempts, and the needle gauge used to influence the level of pain experienced by pediatric patients during IV insertion?

To examine the objective at hand, we constructed two models on the matched subset data for consistency with the first objective. The first model included only the listed five main effects, serving as the base model while the second model introduced interaction effects. The null hypothesis posited that there is no relationship between the dependent variables and the FPS-R Score, while the alternative hypothesis suggested the presence of such a relationship.

Firstly, we conducted linear regression solely on the main effects, adhering to a 5% significance level. Secondly, we investigated how the relationship between one categorical factor and the FPS-R score depends on the value of the second categorical factor. We grouped the age variable into two groups, 5-10 and 11-16 to consider children's development. Additionally, we categorized the number of IV attempts into low and high. Then, we paired all five main effects and generated interaction plots. If the lines intersect, indicating an interaction between the selected two variables, we included the interaction term in the regression model. Once all terms, including the interaction effects, were incorporated, we performed linear regression again.

Secondary Objective 2: The incidence of side effects due to Ametop or the vapocoolant spray.

4

Study Design and Participants: All participants initially included in the study were considered in the analysis, even those who withdrew before its conclusion, to consider patients who withdrew due to experiencing side effects and preserving the comprehensive nature of the data. Thus, all 240 participants of the study are included in this objective's analysis.

Side Effects Categorization: Categories of observed side effects were extracted from the 'Observed side effects' column, where side effects of patients ranged from "Slight redness of skin from Ametop" to "quite red & itchy skin from Ametop. Underlying eczema." These reports were categorized into distinct groups such as the individual side effect "Redness" and the combined side effect "Redness + Itching" to facilitate the quantitative analysis of side effects.

Result

Main Objective: Does the combination of Ametop with a vapocoolant spray significantly reduce the discomfort experienced by pediatric patients during IV insertions, compared to the application of Ametop alone?

Using the entire dataset, the result of the Welch Two Sample T-Test yields a p-value of 0.02258. This suggests that we can reject the null hypothesis at a 5% significance level, indicating that the mean pain level is not the same for both the control and study groups. This supports the idea that the addition of vapocoolant spray may reduce discomfort in pediatric patients.

However, when conducting the t-test on the subset of data where the FPS-R scores and observer's scores align, we obtain a p-value of 0.0594. This suggests that we cannot reject the null hypothesis at a 5% significance level, indicating that the mean pain level is the same for both the control and study groups, and we cannot conclude that the vapocoolant spray reduces discomfort.

Secondary Objective 1: Do age, gender, the practitioner's role, the number of IV attempts, and the needle gauge used to influence the level of pain experienced by pediatric patients during IV insertion?

Base Model:

Age: The coefficient for Age is -0.11476, which is statistically significant ($p < 0.05$). This implies that for each additional year of age, the FPS-R score decreases by approximately 0.115 points, suggesting that older children experience slightly less discomfort during IV insertions.

Number of IV Attempts: This coefficient is positive and significant ($p < 0.05$), indicating that an increased number of IV attempts is associated with a higher FPS-R score, suggesting greater discomfort with more attempts.

Out of 5 main effects (age, gender, the practitioner's role, the number of IV attempts, and the needle gauge) in the base model, only the above two are significant.

Interaction Model:

There are several significant interactions gathered from the interaction plot: Sex and Age, Sex and Needle gauge, Sex and Practitioner, Age and Needle gauge, Age and Practitioner, Number of IV attempts and Practitioner. All these interactions and the five main effects are included in this more comprehensive regression model. (Interaction plots attached in appendix)

Age and Practitioner Interaction: The interaction between age and practitioner type is also significant ($p < 0.05$) with a negative coefficient (-0.412774). This indicates that the effect of a resident performing the procedure has a differential impact based on the patient's age. Specifically, the increase in discomfort associated with a resident performing the IV insertion diminishes as the patient gets older.

Practitioner: There was a significant positive coefficient for practitioners with resident status ($p < 0.05$), suggesting that when residents instead of anesthetists performed the IV insertion, there was an increase in the FPS-R score, which could be interpreted as an increase in perceived discomfort.

These results suggest that the discomfort during IV insertions is influenced not just by the number of attempts or the age of the patient but also by the experience level of the practitioner, and this effect varies with the age of the patient. It highlights the importance of considering these interactions when planning and training for IV insertions to optimize patient comfort.

Secondary Objective 2: The incidence of side effects due to Ametop or the vapocoolant spray.

Overall Tolerance

Observed Side Effects
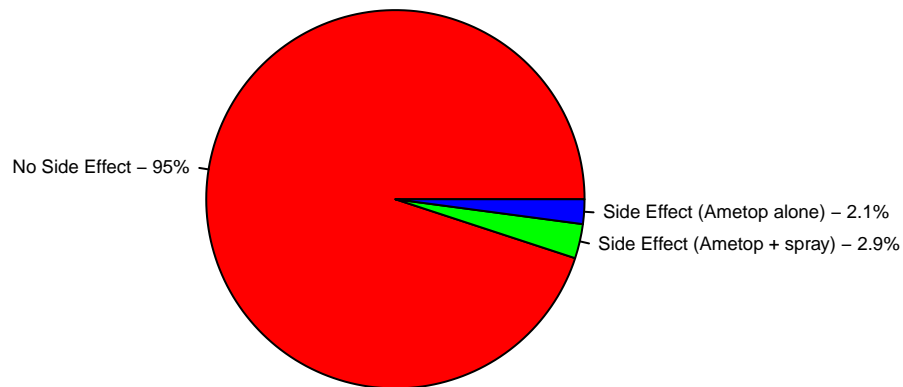
Figure 3: Observed Side Effects

As illustrated in Figure 3, the vast majority of the study participants, 95%, reported no side effects, indicating a high level of tolerance for both treatments. Within the 5% of patients experiencing side effects, only 2.1% were attributed to the use of Ametop alone, whereas the incidence of side effects from the combination treatment was higher, at 2.9%.

Side Effects in Ametop Alone Group

```
##                                             Description        Categories
## 1                  Slight redness of skin from ametop            Redness
## 2                    vey mild pink and pruritic skin.   Redness, Itching
## 3                       pink skin and very mild itching          Redness
## 4 quite red & itchy skin from ametop. Underlying eczema. Redness, Itching
## 5                       redness & raised area under ametop       Redness
```

6
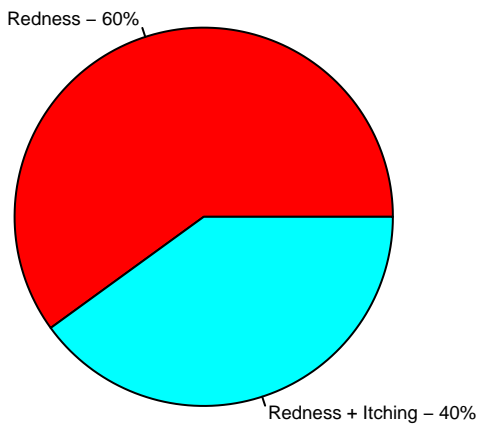
Combination Side Effects Distribution – Ametop alone

Redness – 60%

Redness + Itching – 40%

Figure 4: Combination Side Effects Distribution - Ametop Alone

Individual Side Effects Distribution – Ametop alone
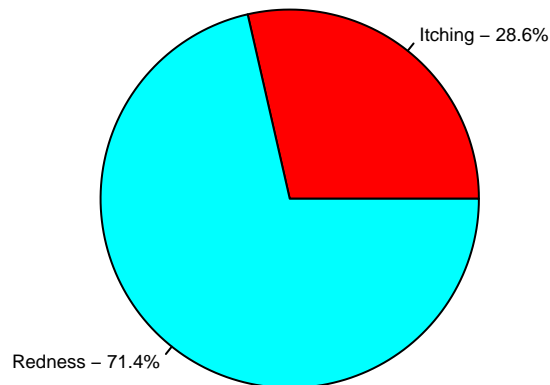
Itching – 28.6%

Redness – 71.4%

Figure 5: Individual Side Effects Distribution - Ametop Alone

Figure 4 and Figure 5 present the distribution of side effects among patients treated with Ametop alone. Figure 4 shows that among those who reported side effects, 60% experienced redness only, while 40% experienced both redness and itching. No other combinations of side effects were reported. This suggests a tendency for side effects to present as redness alone or in conjunction with itching. Figure 5 further highlights that redness is the more prevalent individual side effect, reported by 71.4% of patients experiencing side effects, while itching was reported by 28.6%.

Side Effects in Ametop with Vapocoolant Spray Group

```
##                                     Description       Categories
## 1          Some redness on skin, more than usual         Redness
## 2                  complained of pain on the spray.          Pain
## 3       slightly puffy hands after ametop removal       Puffiness
## 4                            redness from ametop          Redness
## 5 redness at ametop site. Not painful or pruritic.        Redness
## 6          quite itchy from ametop. Redness. Redness, Itching
```

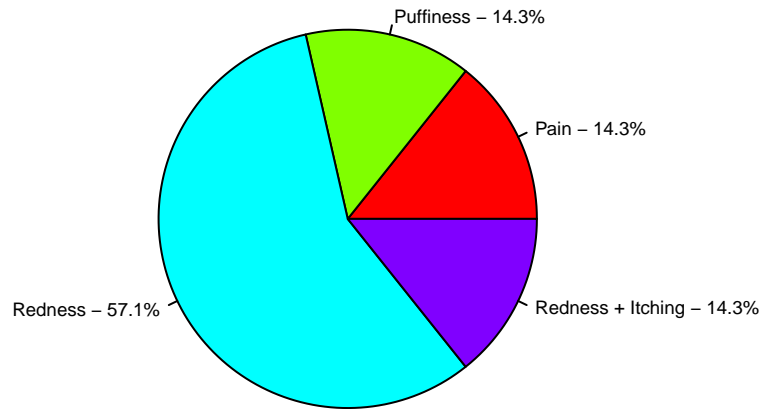## Combination Side Effects Distribution – Ametop + Spray

Figure 6: Combination Side Effects Distribution - Ametop + Spray

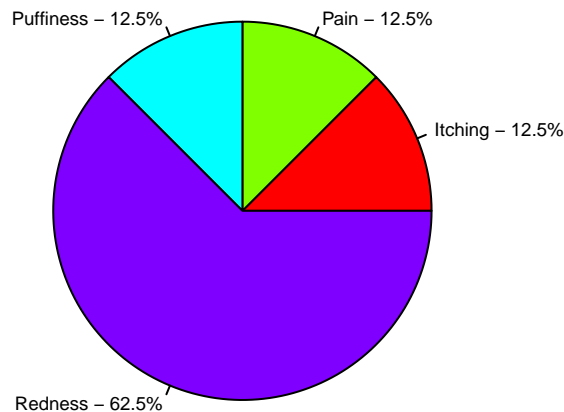## Individual Side Effects Distribution – Ametop + Spray

Figure 7: Individual Side Effects Distribution - Ametop + Spray

For patients treated with both Ametop and a vapocoolant spray, Figure 6 and Figure 7 depict the side effects. Figure 6 reveals a varied response, with redness alone constituting 57.1% of side effects, and pain, puffiness, and redness with itching each accounting for 14.3%. Figure 7 indicates that redness remains the most reported individual side effect at 62.5%, with itching, pain, and puffiness equally reported at 12.5% each.

Conclusion

The conclusion of our main objective, which is to determine whether the addition of vapocoolant spray significantly reduces discomfort experienced by pediatric patients during IV insertions compared to Ametop alone, depends on the dataset used. When analyzing the full dataset, the results suggest that the addition of vapocoolant spray does significantly reduce discomfort. However, when examining the subset of data where the FPS-R scores and observer's scores align, there appears to be no significant difference in discomfort

between the control and study groups. We favor the results from the subset data as it indicates more reliable and higher-quality data. Nonetheless, we acknowledge limitations in how the data is subsetted and plan to discuss with the client in our second meeting potential alternative methods for subsetting the data to improve results without introducing bias or removing too many observations.

Regarding our secondary objective, which is to identify other variables that influence the level of pain experienced by pediatric patients during IV insertion, we conclude that age, number of IV attempts, and practitioner's role are significant factors. It emphasizes the significance of taking these factors into account during IV insertions to alleviate patients' discomfort. A limitation observed in this regression method is that the data does not appear to be exactly normally distributed, as indicated by the QQ plot, which shows a heavy-tailed distribution. However, since the group is focusing on inference rather than prediction, and considering the Central Limit Theorem (CLT) assumption, we will still proceed with using linear regression for our analysis.

The inability to reject the null hypothesis in our main objective could be explained by the significant variables found in the regression model. Age, practitioner roles, and number of IV attempts have been identified as significant in influencing pain levels. Segmenting the data differently, perhaps by excluding certain age groups, number of IV attempts, or practitioner roles, might yield different results. Our next step is to retest the main objective, assessing whether there is a difference in the FPS-R scores between the control and study groups when segmented by age, practitioner role, and number of IV attempts, to see if this impacts the results in any way.

In analyzing the side effects of Ametop and its combination with a vapocoolant spray, it's evident that the majority of patients (95%) experience no side effects, underlining the treatments' high tolerance. Among the 5% that do report side effects, the occurrences are mild, such as redness or itchiness. For those treated with Ametop alone, side effects were predominantly redness, either alone or accompanied by itching, which is valuable information for healthcare providers assessing the drug's tolerability. On the other hand, the combination treatment group presented a more varied range of side effects, with equal instances of itching, pain, and puffiness, in addition to redness. However, we were not sure whether the side effect was a result of the Ametop or the spray for the treatment group. Overall, the observed variations in side effect profiles highlight the significance of personalized patient care and the broad range of reactions suggests that clinicians must exercise careful consideration when prescribing treatments.
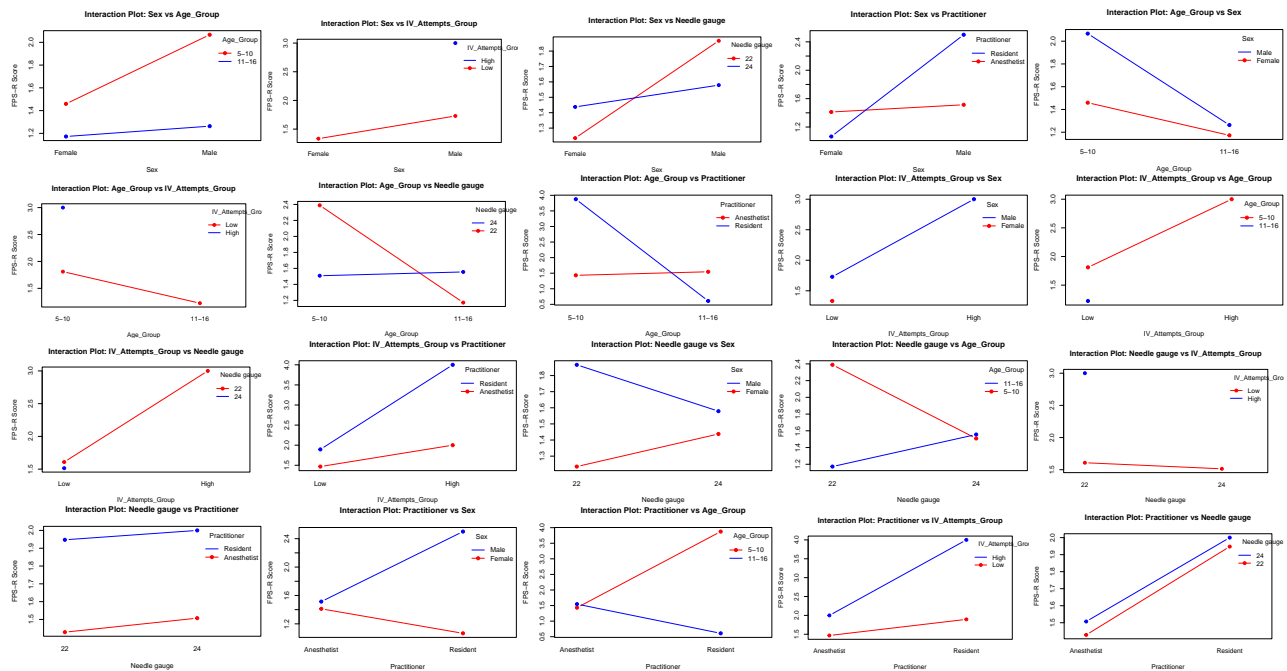
References

Rogers TL, Ostrow CL. The use of EMLA cream to decrease venipuncture pain in children. Journal of Pediatric Nursing. 2004 Feb 1;19(1):33-9.

Appendix

For client

- Interaction plots

[1] "Categorical variables: Sex, Age_Group, IV_Attempts_Group, Needle gauge, Practitioner"
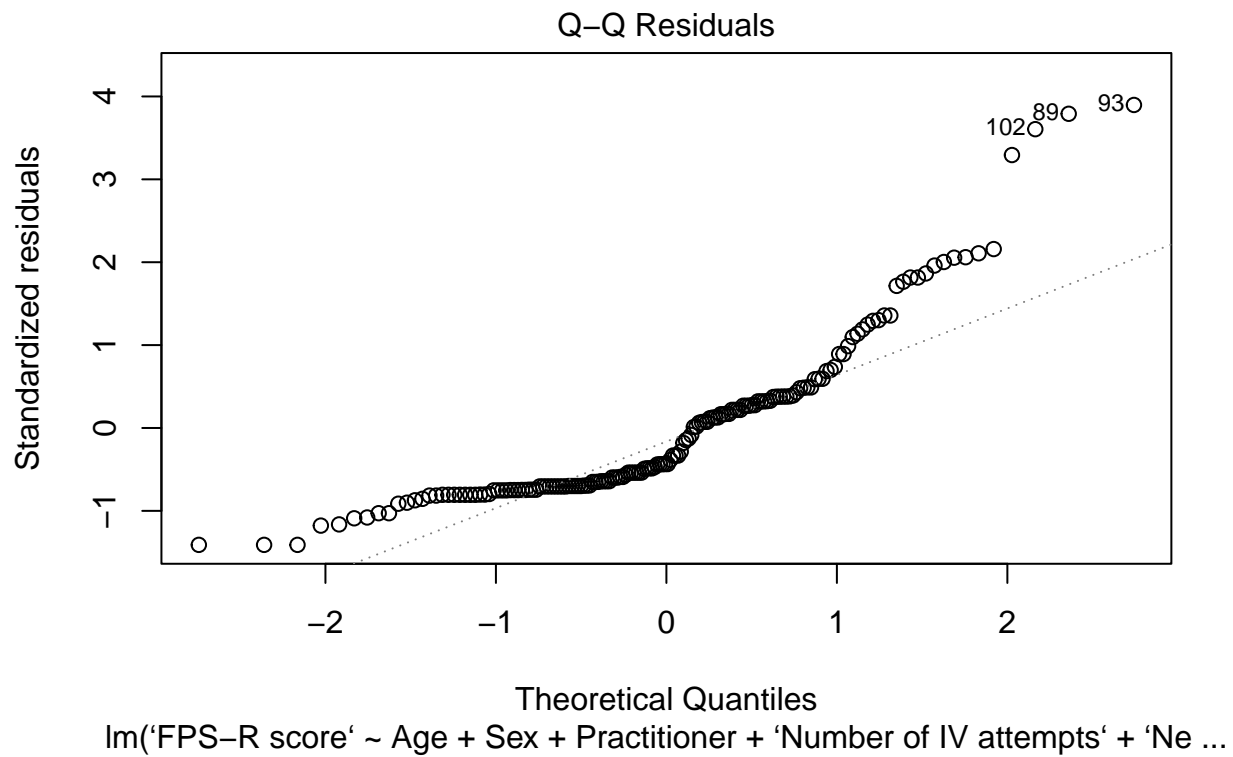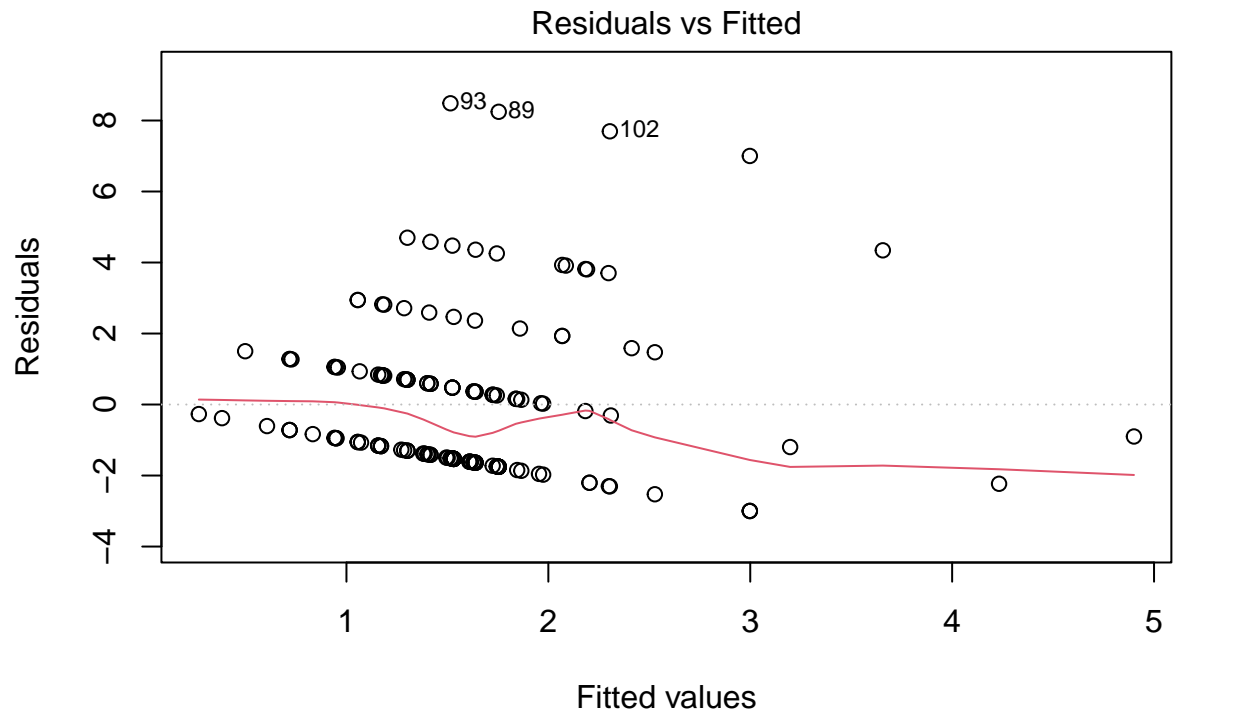

- Linear regression output and QQ plot

```
##
## Call:
## lm(formula = `FPS-R score` ~ Age + Sex + Practitioner + `Number of IV attempts` +
##     `Needle gauge`, data = data_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9980 -1.5312 -0.9209  0.8204  8.4849
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             6.25585    5.92007   1.057   0.2923
## Age                    -0.11476    0.05758  -1.993   0.0480 *
## SexMale                 0.22291    0.35796   0.623   0.5344
## PractitionerResident    0.43860    0.45404   0.966   0.3355
## `Number of IV attempts`  1.24388    0.54460   2.284   0.0237 *
## `Needle gauge`         -0.22478    0.23880  -0.941   0.3480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.201 on 158 degrees of freedom
## Multiple R-squared:  0.07538,    Adjusted R-squared:  0.04612
## F-statistic: 2.576 on 5 and 158 DF,  p-value: 0.02855
##
```
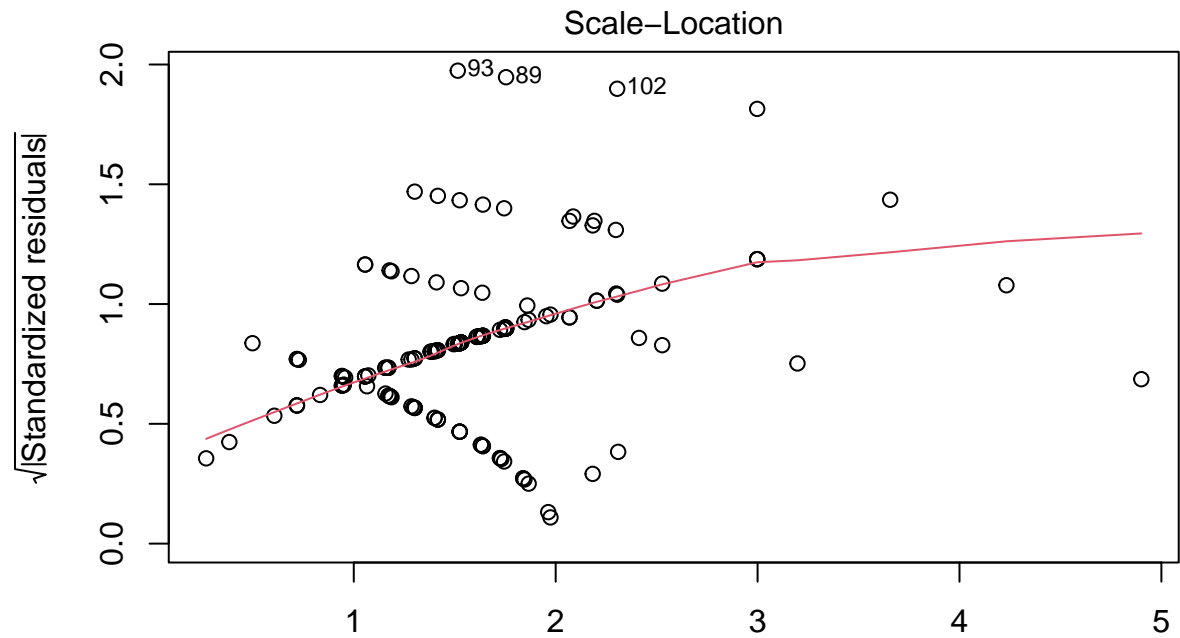
```
## Call:
## lm(formula = `FPS-R score` ~ Age + Sex + Practitioner + `Number of IV attempts` +
##     `Needle gauge` + Sex * Age + Sex * `Needle gauge` + Sex *
##     Practitioner + Age * `Needle gauge` + Age * Practitioner +
##     `Number of IV attempts` * Practitioner, data = data_subset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5451 -1.3624 -0.8256  0.6689  8.6001
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                               1.45661   17.65037   0.083
## Age                                      -0.36148    1.51844  -0.238
## SexMale                                   5.26317   11.89520   0.442
## PractitionerResident                      5.47138    2.36973   2.309
## `Number of IV attempts`                   1.15804    0.65252   1.775
## `Needle gauge`                           -0.06282    0.74856  -0.084
## Age:SexMale                              -0.05833    0.11586  -0.503
## SexMale:`Needle gauge`                   -0.20319    0.48571  -0.418
## SexMale:PractitionerResident              0.40589    0.96106   0.422
## Age:`Needle gauge`                        0.01671    0.06584   0.254
## Age:PractitionerResident                 -0.41277    0.14123  -2.923
## PractitionerResident:`Number of IV attempts` -0.56818    1.16120  -0.489
##                                          Pr(>|t|)
## (Intercept)                                0.9343
## Age                                        0.8122
## SexMale                                    0.6588
## PractitionerResident                       0.0223 *
## `Number of IV attempts`                    0.0779 .
## `Needle gauge`                             0.9332
## Age:SexMale                                0.6154
## SexMale:`Needle gauge`                     0.6763
## SexMale:PractitionerResident               0.6734
## Age:`Needle gauge`                         0.8000
## Age:PractitionerResident                   0.0040 **
## PractitionerResident:`Number of IV attempts`  0.6253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.155 on 152 degrees of freedom
## Multiple R-squared:  0.1474, Adjusted R-squared:  0.08571
## F-statistic: 2.389 on 11 and 152 DF,  p-value: 0.009291
```
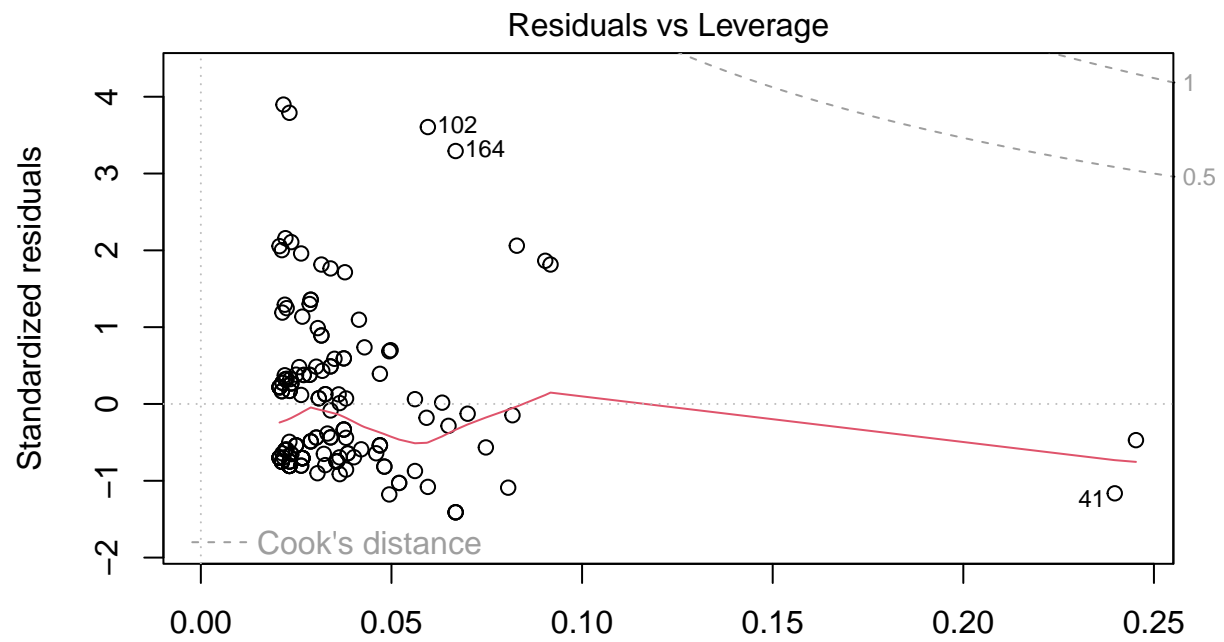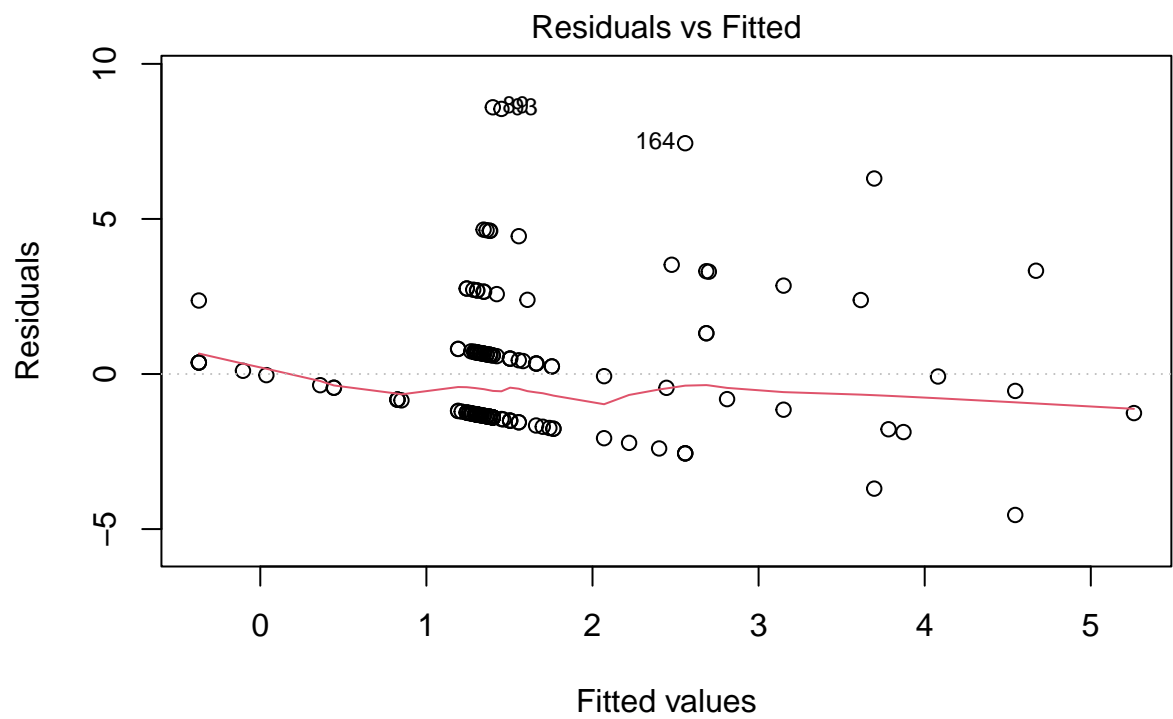
Print linear model without interactions

Residuals vs Fitted

Fitted values
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Q–Q Residuals

Theoretical Quantiles
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Scale–Location

√|Standardized residuals|

○93 ○89   ○102

Fitted values
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Residuals vs Leverage

Standardized residuals

○102
○164

1

0.5

- - - Cook's distance

41○

Leverage
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Print linear model with interactions

Residuals vs Fitted

Fitted values
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Q−Q Residuals

Theoretical Quantiles
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Scale−Location

√|Standardized residuals|

164

Fitted values
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Residuals vs Leverage

Standardized residuals

164

102

69

Cook's distance
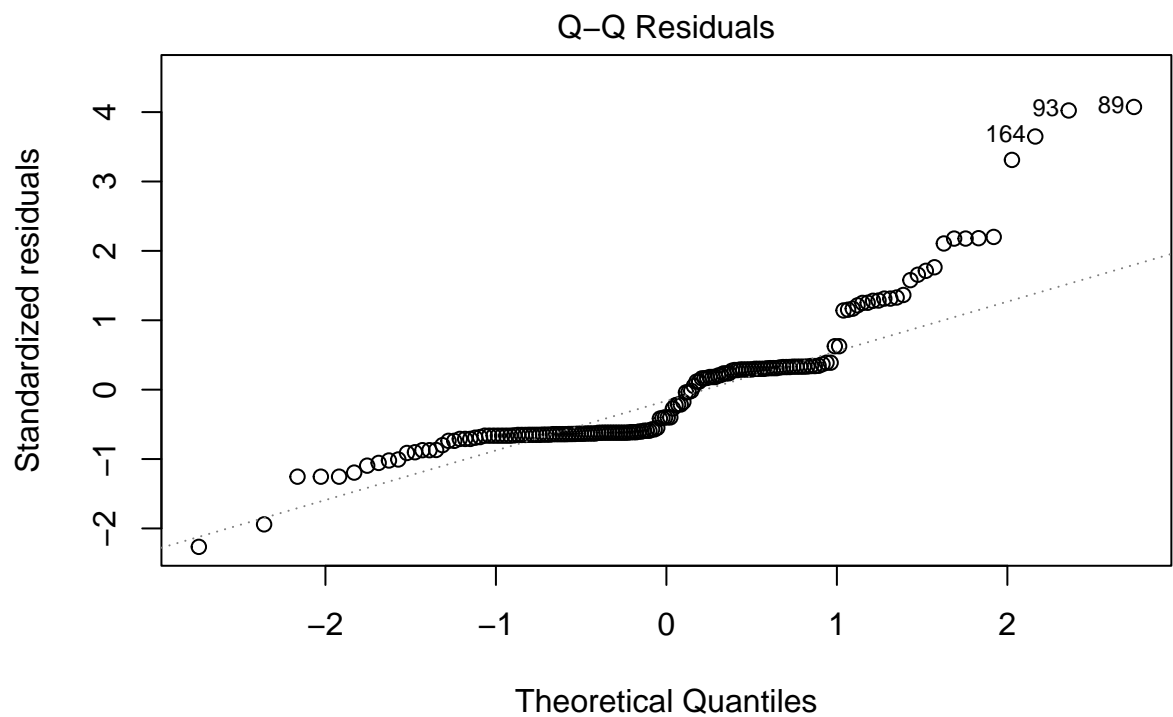
Leverage
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...
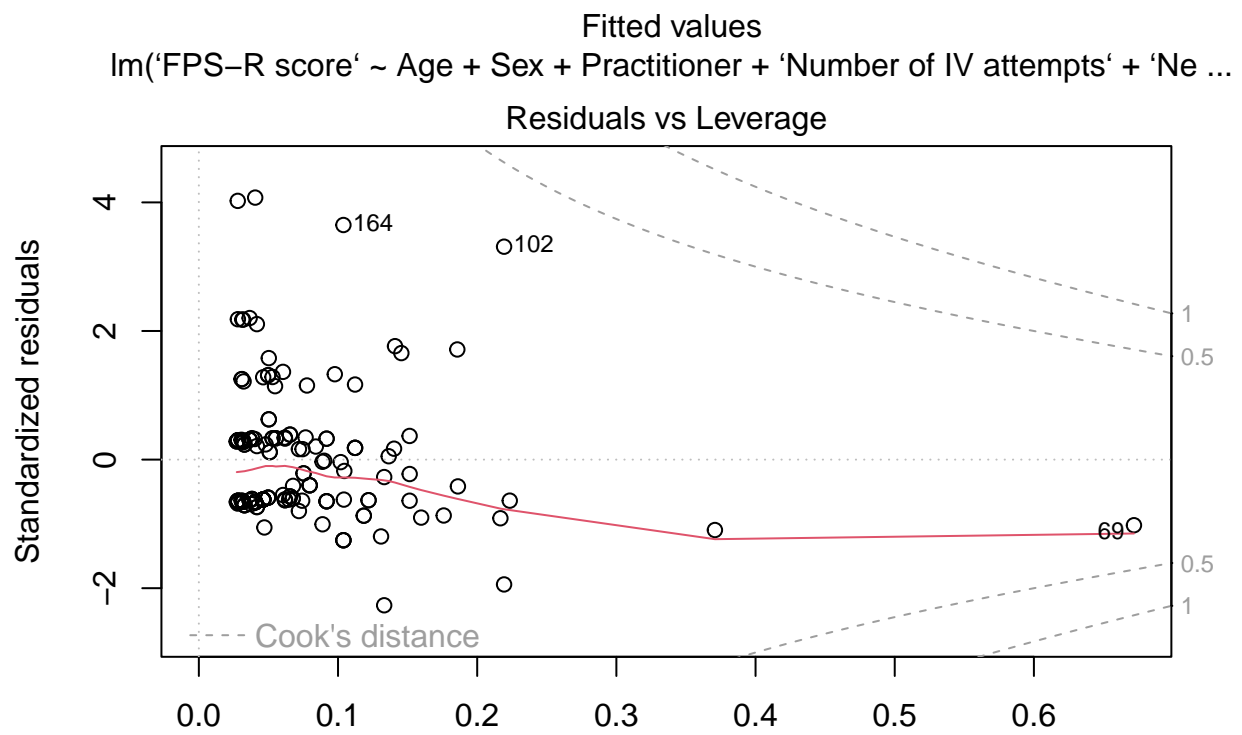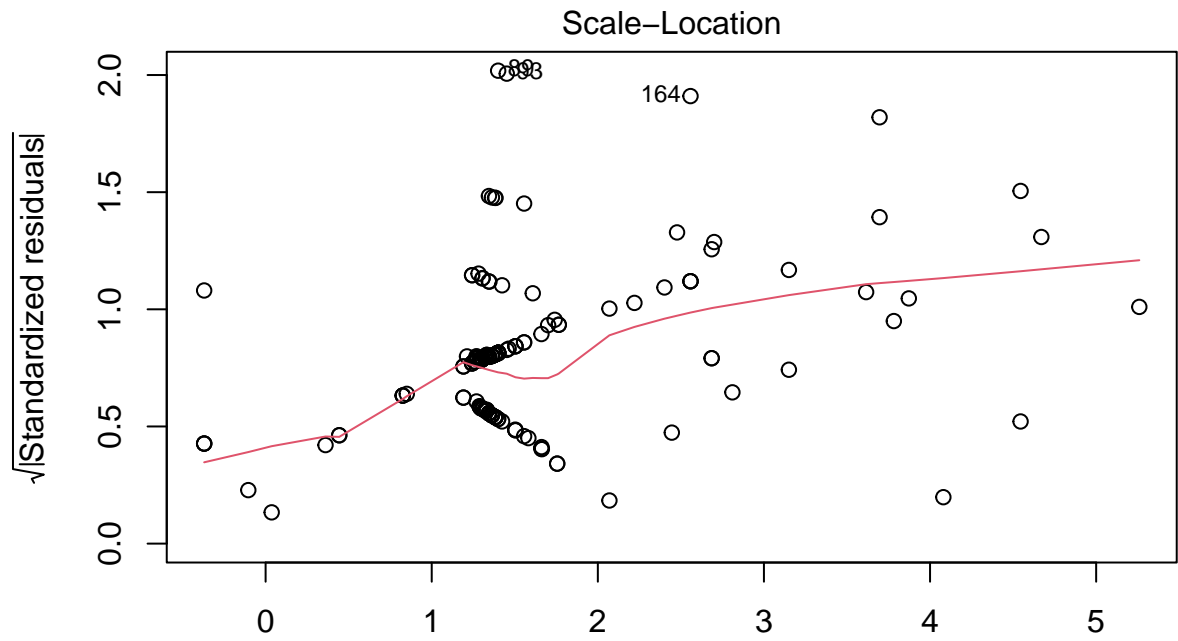
For mentor

Data cleaning

```r
# Read data
file_path <- "../data/Ametop-PE data(new-Alex)(1control 2study).xlsx"
original_data <- read_excel(file_path)

# Save original data for side effects analysis
data <- original_data

# Remove withdrawn patients
data <- data[data$`Withdrawn after consent?` != "Yes", ]

# Remove unnecessary variables
data <- select(data,
               -`Elective surgery requiring IV?`,
               -`ASA I or II?`,
               -`Aged 5-16?`,
               -`Any allergies to Ametop, Pain Ease, or Tagederm adhesive?`,
               -`Ametop placed at least 30min before estimated IV start time?`,
               -`Receiving sedative pre-medication / anxiolytics?`,
               -`Needle phobia?`,
               -`Planned inhalation induction?`,
               -`Developmental delay or unable to interpret FPS-R?`,
               -`Time Ametop applied`,
               -`Time Ametop removed`,
               -`Time of skin puncture`,
               -`Record ID`,
               -`Complete?`,
               -`Withdrawn after consent?`,
               -`Reason for withdrawal`,
               -`Other reason for withdrawal`)
```

```r
# Check for missing values in all columns except 'Notes' and 'Observed side effects'
missing_values_summary <- data %>%
  select(-Notes, -`Observed side effects`) %>%
  summarise_all(~sum(is.na(.)))

# No missing values found
print(missing_values_summary)
```

```
## # A tibble: 1 x 11
##   Randomization   Age   Sex Practitioner `Duration of Ametop`
##           <int> <int> <int>        <int>                <int>
## 1             0     0     0            0                    0
## # i 6 more variables: `Time between Ametop and skin puncture` <int>,
## #   `Reaction to skin puncture` <int>, `Number of IV attempts` <int>,
## #   `FPS-R score` <int>, `Needle gauge` <int>,
## #   `Any side effects observed?` <int>
```

```r
# Shorten the data
data$`Reaction to skin puncture` <- gsub("^(Slight|Severe).*", "\\1 Pain",
                                         data$`Reaction to skin puncture`)

# Verify the changes
```

```
table(data$`Reaction to skin puncture`)
```

```
##
##      None Severe Pain Slight Pain
##       158          5          60
```

Data preprocessing

```
# Create new variables for the dataset

# Create 'Observer Number'
data$`Observer Number` <- case_when(
  data$`Reaction to skin puncture` == "None" ~ 1,
  data$`Reaction to skin puncture` == "Slight Pain" ~ 2,
  data$`Reaction to skin puncture` == "Severe Pain" ~ 3
)

# Create 'FPS-R Number'
data$`FPS-R Number` <- case_when(
  data$`FPS-R score` %in% c(0, 2) ~ 1,
  data$`FPS-R score` %in% c(4, 6) ~ 2,
  data$`FPS-R score` %in% c(8, 10) ~ 3
)

# Create 'Match Responses'
data$`Match Responses` <- ifelse(data$`Observer Number` == data$`FPS-R Number`, "Yes", "No")

# Create data subset for match data
data_subset <- data[data$`Match Responses` == 'Yes',]

# Verify data
head(data)
```

```
## # A tibble: 6 x 16
##   Randomization   Age Sex    Practitioner `Duration of Ametop`
##           <dbl> <dbl> <chr>  <chr>                       <dbl>
## 1             2     6 Female Resident                       50
## 2             1     9 Female Anesthetist                    60
## 3             2     7 Male   Resident                       48
## 4             1     8 Female Anesthetist                    37
## 5             1     9 Male   Resident                       60
## 6             1     6 Female Anesthetist                    34
## # i 11 more variables: `Time between Ametop and skin puncture` <dbl>,
## #   `Reaction to skin puncture` <chr>, `Number of IV attempts` <dbl>,
## #   `FPS-R score` <dbl>, `Needle gauge` <dbl>,
## #   `Any side effects observed?` <chr>, `Observed side effects` <chr>,
## #   Notes <chr>, `Observer Number` <dbl>, `FPS-R Number` <dbl>,
## #   `Match Responses` <chr>
```

Objective 1

```
# Histogram
plot1 <- data %>% filter(Randomization == 1) %>%
  ggplot(aes(`FPS-R score`)) +
```

```r
  geom_histogram() +
  labs(title = "Group 1 Histogram")

plot2 <- data %>% filter(Randomization == 2) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_histogram() +
  labs(title = "Group 2 Histogram")

# Boxplot
plot3 <- data %>% filter(Randomization == 1) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Full Data Group 1 Boxplot")

plot4 <- data %>% filter(Randomization == 2) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Full Data Group 2 Boxplot")

# Boxplot
plot5 <- data %>% filter(Randomization == 1, `Match Responses` == "Yes") %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Match Subset Data Group 1 Boxplot")

plot6 <- data %>% filter(Randomization == 2, `Match Responses` == "Yes") %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Match Subset Data Group 2 Boxplot")

grid.arrange(plot1,plot2,plot3,plot4,plot5,plot6, ncol = 2, nrow = 3)
```
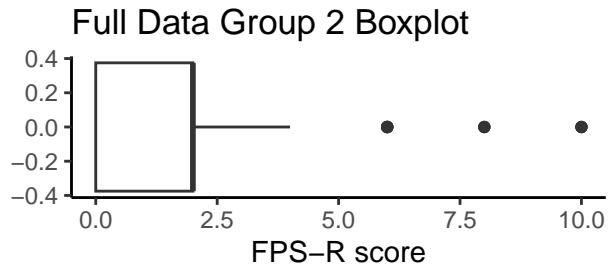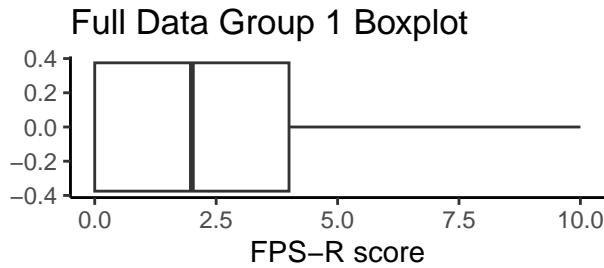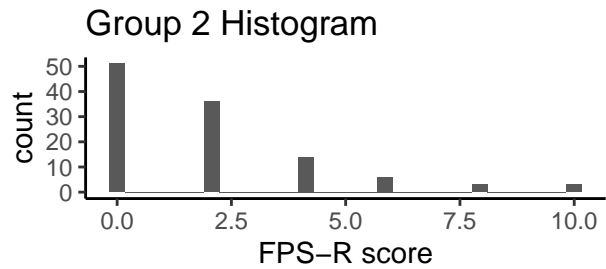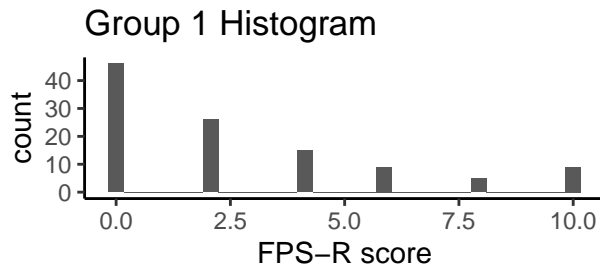
Histrograms and boxplots

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Group 1 Histogram

Group 2 Histogram

Full Data Group 1 Boxplot

Full Data Group 2 Boxplot

Match Subset Data Group 1 Boxplot

Match Subset Data Group 2 Boxplot
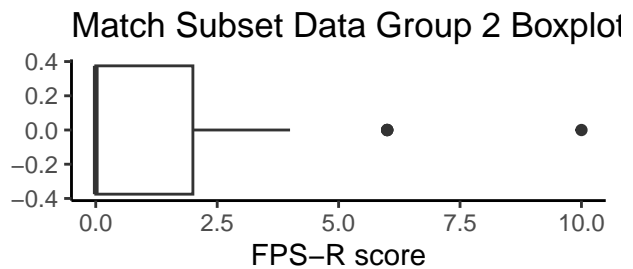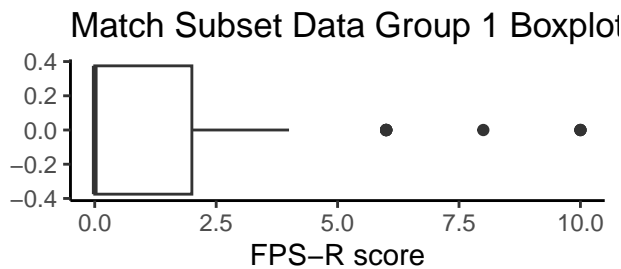
```r
# Create a table to count matches
table <- table(data$`Observer Number`, data$`FPS-R Number`)

# Name the rows and columns to reflect the desired format
rownames(table) <- c("1", "2", "3")
colnames(table) <- c("1", "2", "3")

# Output the 3x3 table
cat("Contingency Table of FPS-R Number vs. Observer Number\n")
```

Table

```
## Contingency Table of FPS-R Number vs. Observer Number
```

```r
cat("Rows: FPS-R Number\n")
```

```
## Rows: FPS-R Number
```

```r
cat("Columns: Observer Number\n\n")
```

```
## Columns: Observer Number
```

```r
print(table)
```

```
##
##      1   2   3
##   1 135  20   3
##   2  24  24  12
##   3   0   0   5
```

```r
# Perform t-test on full data
t_test_full <- t.test(data[data$Randomization == 2, ]$`FPS-R score`,
                       data[data$Randomization == 1, ]$`FPS-R score`,
                       alternative = c("less"))
print(t_test_full)
```

T-tests

```
##
##  Welch Two Sample t-test
##
## data:  data[data$Randomization == 2, ]$`FPS-R score` and data[data$Randomization == 1, ]$`FPS-R score`
## t = -2.0155, df = 204.03, p-value = 0.02258
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.1372509
## sample estimates:
## mean of x mean of y
##  1.929204  2.690909
```

```r
# Perform t-test on subset of data
t_test_subset <- t.test(data_subset[data_subset$Randomization == 2, ]$`FPS-R score`,
                         data_subset[data_subset$Randomization == 1, ]$`FPS-R score`,
                         alternative = c("less"))
print(t_test_subset)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_subset[data_subset$Randomization == 2, ]$`FPS-R score` and data_subset[data_subset$Rando
## t = -1.5712, df = 118.53, p-value = 0.0594
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf 0.03207875
## sample estimates:
## mean of x mean of y
##  1.333333  1.915493
```

```r
# Conclusion: We should only reject the null hypothesis based on the full dataset.
# If we consider the subset of data with 164 observations (after excluding 59
# from the original 223), we do not reject the null hypothesis.
```

Objective 2

```r
# Only use match subset data

# Convert continuous variables to categorical factors
data_subset$Age_Group <- cut(data_subset$Age,
                             breaks = 2, labels = c("5-10", "11-16"))
data_subset$IV_Attempts_Group <- cut(data_subset$`Number of IV attempts`,
                                     breaks = 2, labels = c("Low", "High"))

# Make sure all categorical variables are treated as factors
data_subset$`Needle gauge` <- as.factor(data_subset$`Needle gauge`)
```

```r
data_subset$Sex <- as.factor(data_subset$Sex)
data_subset$Practitioner <- as.factor(data_subset$Practitioner)

# List of categorical variables
categorical_vars <- c("Sex", "Age_Group", "IV_Attempts_Group", "Needle gauge", "Practitioner")
print(paste0("Categorical variables: ", paste(categorical_vars, collapse = ", ")))
```
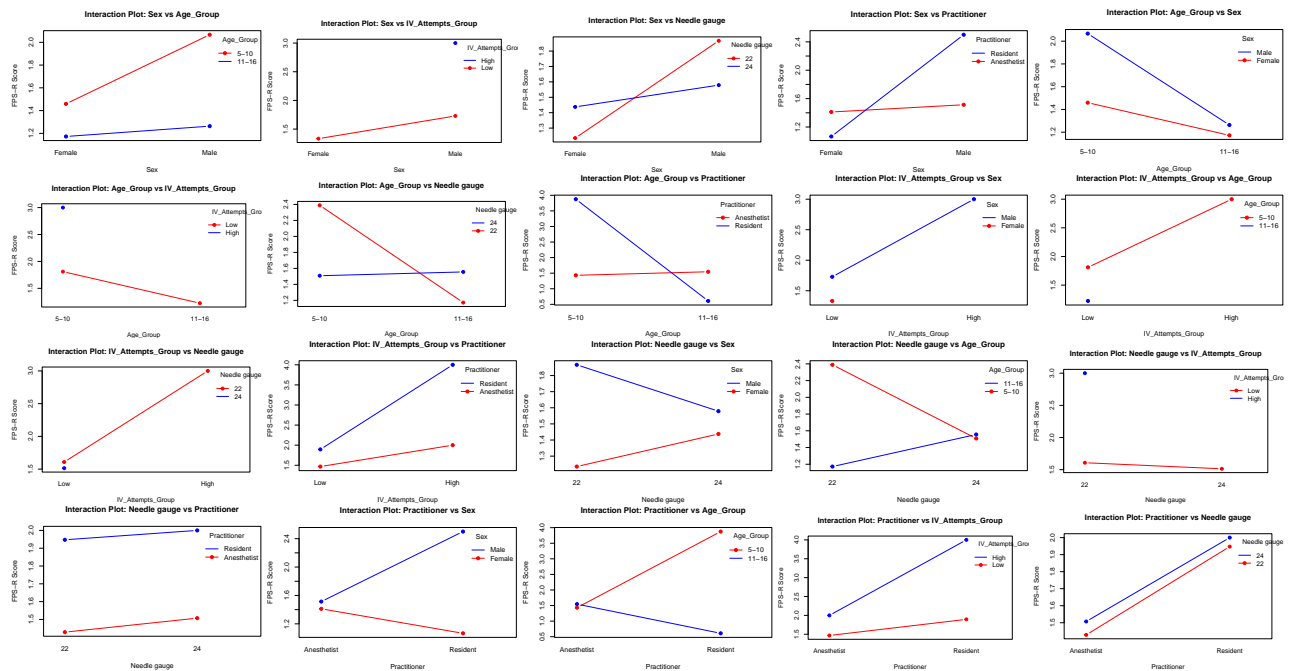
Interaction plots

```
## [1] "Categorical variables: Sex, Age_Group, IV_Attempts_Group, Needle gauge, Practitioner"
```

```r
# Create interaction plots for all pairs of categorical variables
for (x_var in categorical_vars) {
  for (trace_var in categorical_vars) {
    if (x_var != trace_var) {  # Avoid plotting a variable against itself
      interaction.plot(x.factor = data_subset[[x_var]],
                       trace.factor = data_subset[[trace_var]],
                       response = data_subset$`FPS-R score`,
                       type = "b",
                       pch = 19,
                       xlab = x_var,
                       ylab = "FPS-R Score",
                       trace.label = trace_var,
                       col = c("red", "blue"),
                       lty = 1,
                       legend = TRUE,
                       main = paste("Interaction Plot:", x_var, "vs", trace_var))
    }
  }
}
```



```r
# Significant interactions: Sex * Age + Sex * `Needle gauge` + Sex * Practitioner
# + Age * `Needle gauge` + Age * Practitioner + `Number of IV attempts` * Practitioner
```

```r
# General function to create plots
plot_data <- function(data, x_var, y_var = "FPS-R score", x_label = "", y_label = "FPS-R score") {
  p <- data %>%
    ggplot(aes(x = !!sym(x_var), y = !!sym(y_var))) +
    geom_point(size = 2) +
    scale_y_continuous(breaks = seq(0, 10, by = 1)) +
    labs(x = x_label, y = y_label) +
    ggtitle(paste("Distribution of", y_label, "by", x_label))

  print(p)
}
```
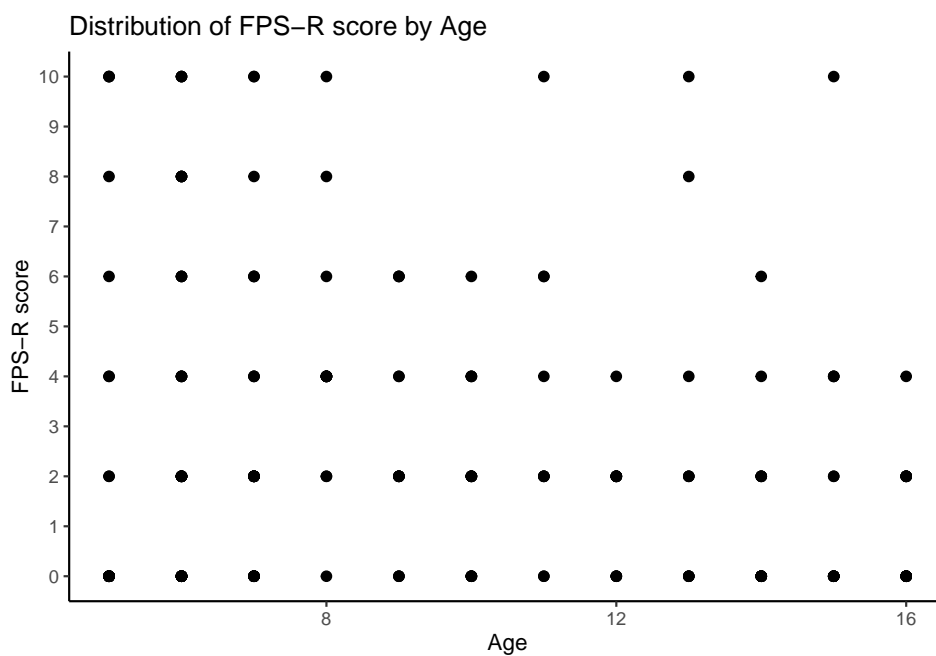
```r
# Plot for Age
plot_data(data, "Age", x_label = "Age")
```
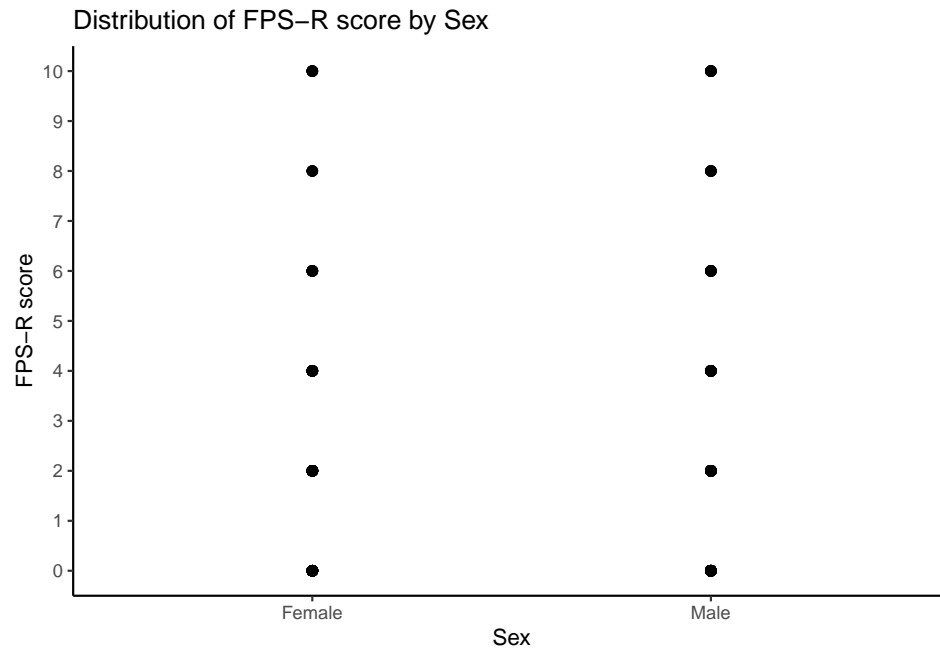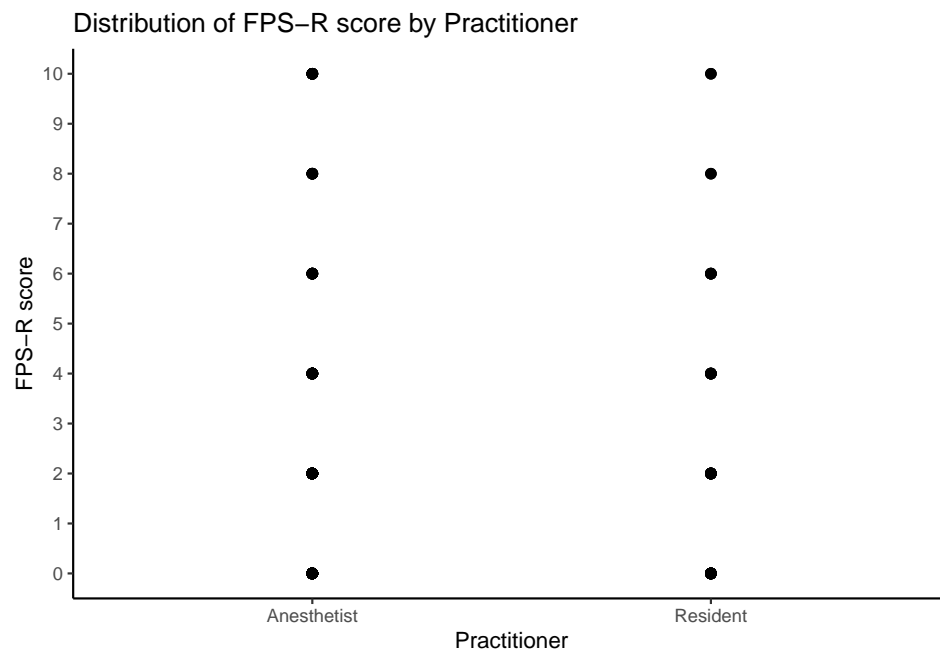
Histograms for FPS-R score

Distribution of FPS−R score by Age

```r
# Plot for Sex
plot_data(data, "Sex", x_label = "Sex")
```

Distribution of FPS−R score by Sex


```
# Plot for Practitioner
plot_data(data, "Practitioner", x_label = "Practitioner")
```
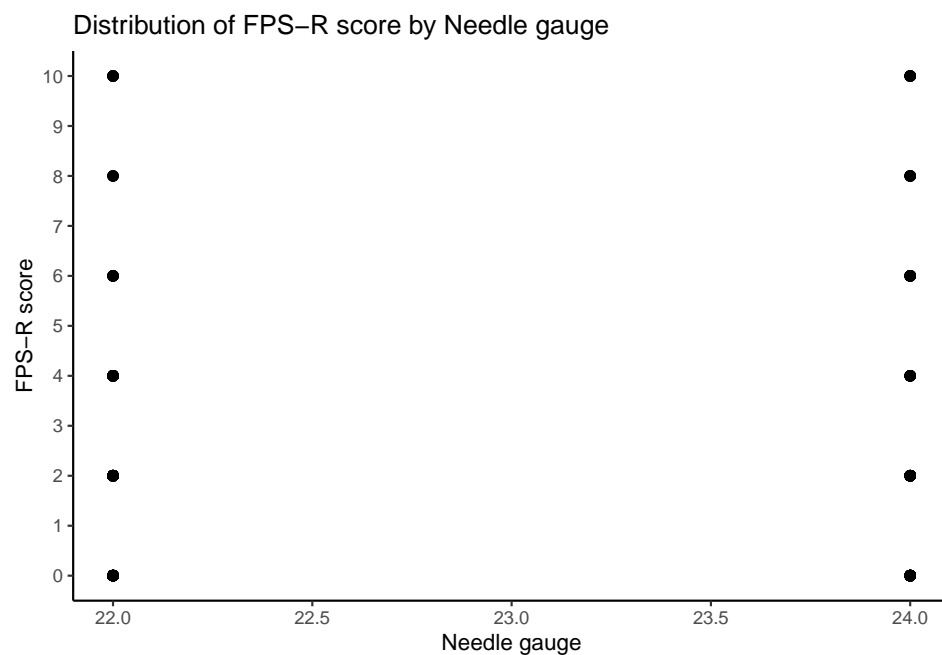
Distribution of FPS−R score by Practitioner


```
# Plot for Number of IV attempts
plot_data(data, "Number of IV attempts", x_label = "Number of IV attempts")
```

Distribution of FPS–R score by Number of IV attempts
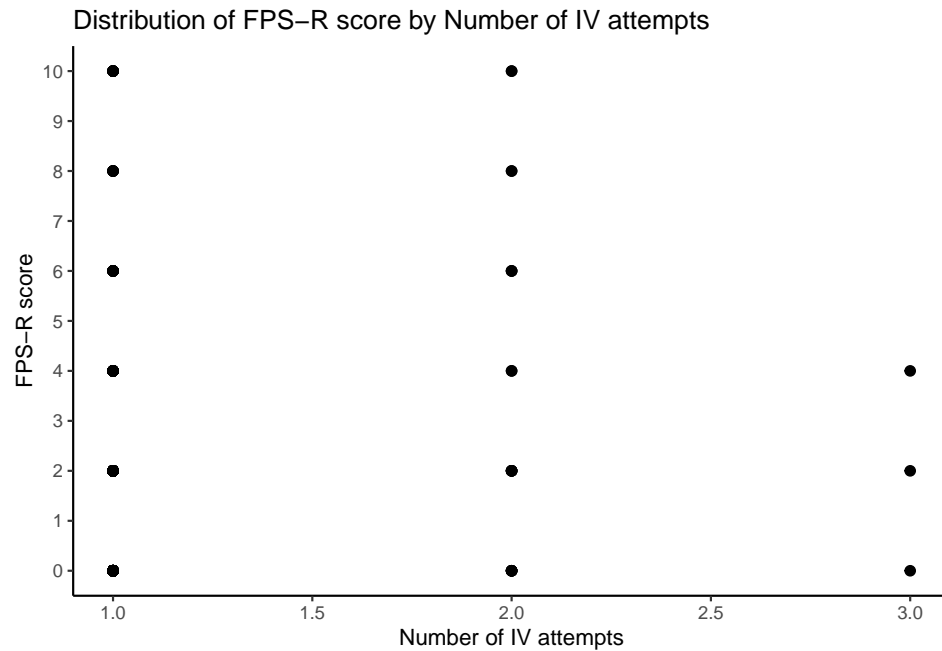

```r
# Plot for Needle gauge
plot_data(data, "Needle gauge", x_label = "Needle gauge")
```

Distribution of FPS–R score by Needle gauge


```r
# The younger the age, the more concentrated the distribution of pain scores.
# The fewer the number of IV injections, the more concentrated the high pain scores are.
```

```r
data_subset <- data[data$`Match Responses` == 'Yes',]

lm_model_without_interactions <- lm(`FPS-R score` ~ Age + Sex + Practitioner +
                `Number of IV attempts` + `Needle gauge`, data = data_subset)
summary(lm_model_without_interactions)
```

Linear models for FPS-R score

```
## 
## Call:
## lm(formula = `FPS-R score` ~ Age + Sex + Practitioner + `Number of IV attempts` +
##     `Needle gauge`, data = data_subset)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9980 -1.5312 -0.9209  0.8204  8.4849
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              6.25585    5.92007   1.057   0.2923
## Age                     -0.11476    0.05758  -1.993   0.0480 *
## SexMale                  0.22291    0.35796   0.623   0.5344
## PractitionerResident     0.43860    0.45404   0.966   0.3355
## `Number of IV attempts`  1.24388    0.54460   2.284   0.0237 *
## `Needle gauge`          -0.22478    0.23880  -0.941   0.3480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.201 on 158 degrees of freedom
## Multiple R-squared:  0.07538,    Adjusted R-squared:  0.04612
## F-statistic: 2.576 on 5 and 158 DF,  p-value: 0.02855
```
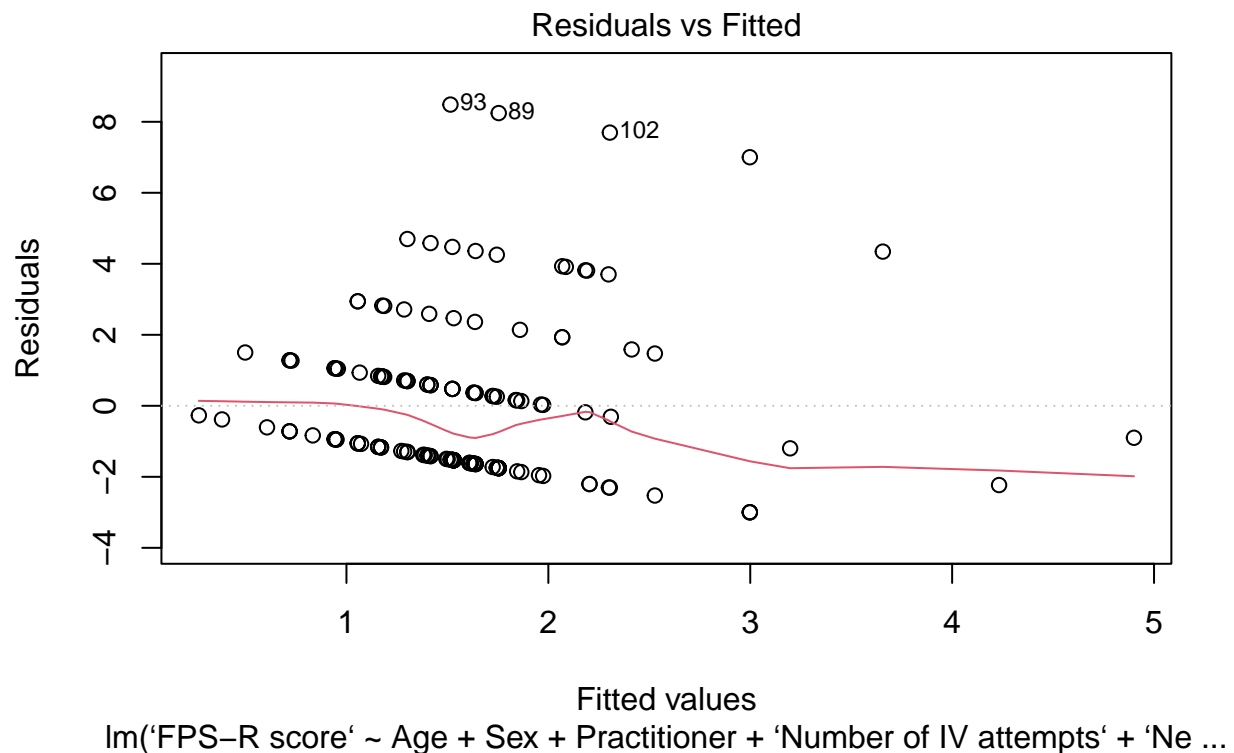
```r
lm_model_with_interactions <- lm(`FPS-R score` ~ Age + Sex + Practitioner + `Number of IV attempts`
                + `Needle gauge` + Sex * Age + Sex * `Needle gauge` + Sex * Practitioner
                + Age * `Needle gauge` + Age * Practitioner + `Number of IV attempts` * Practitioner,
                data = data_subset)
summary(lm_model_with_interactions)
```
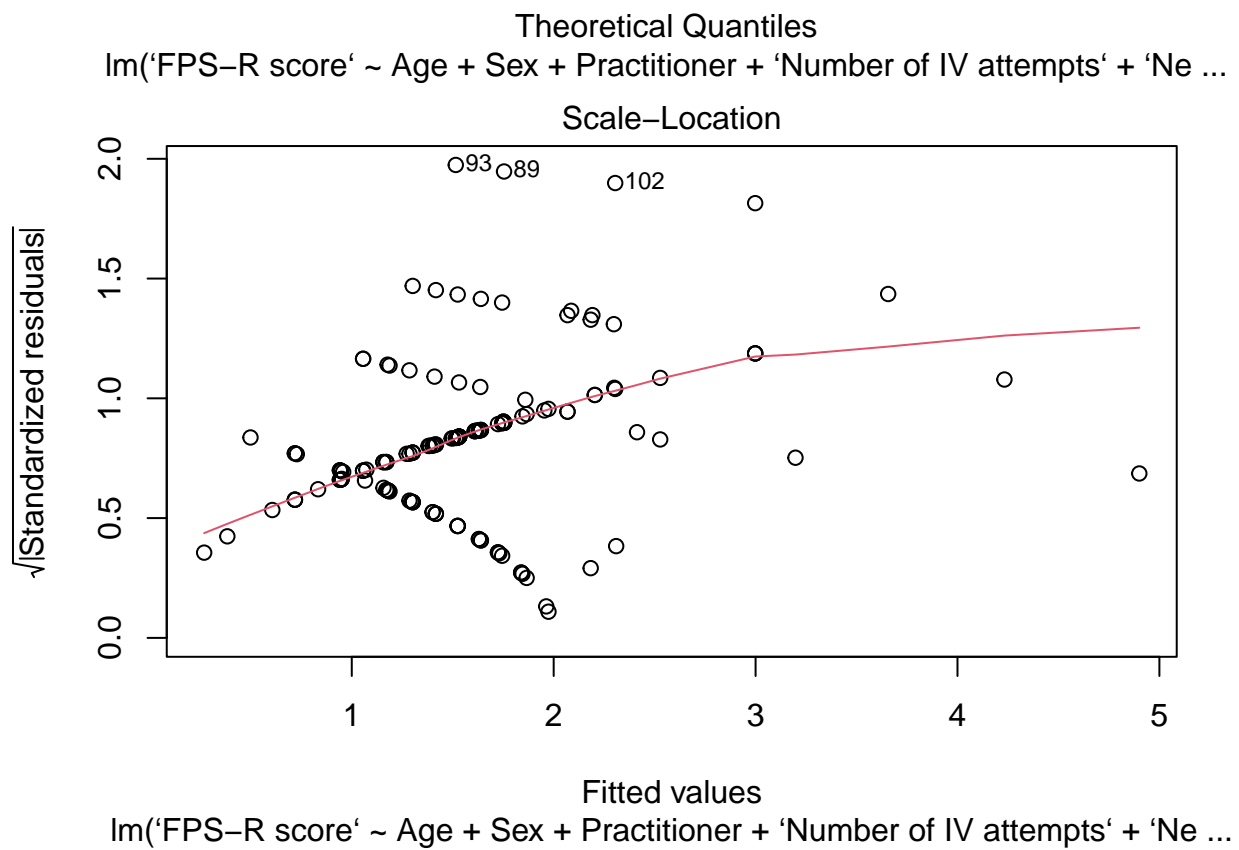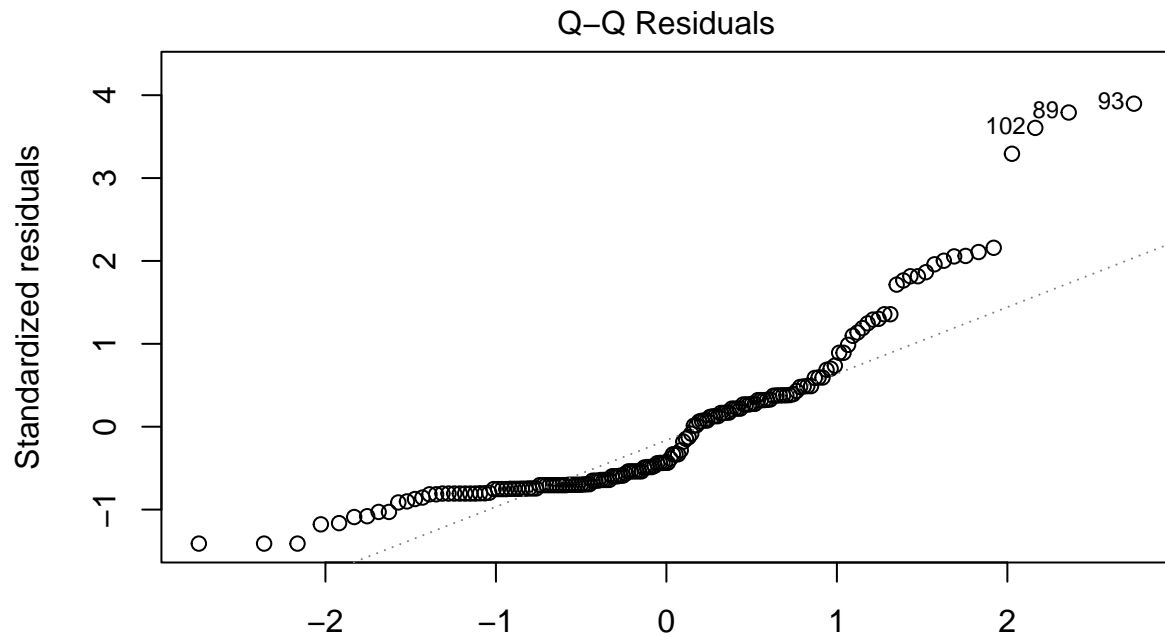
```
## 
## Call:
## lm(formula = `FPS-R score` ~ Age + Sex + Practitioner + `Number of IV attempts` +
##     `Needle gauge` + Sex * Age + Sex * `Needle gauge` + Sex *
##     Practitioner + Age * `Needle gauge` + Age * Practitioner +
##     `Number of IV attempts` * Practitioner, data = data_subset)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5451 -1.3624 -0.8256  0.6689  8.6001
## 
## Coefficients:
##                                  Estimate Std. Error t value
## (Intercept)                       1.45661   17.65037   0.083
## Age                              -0.36148    1.51844  -0.238
## SexMale                           5.26317   11.89520   0.442
## PractitionerResident              5.47138    2.36973   2.309
## `Number of IV attempts`           1.15804    0.65252   1.775
## `Needle gauge`                   -0.06282    0.74856  -0.084
## Age:SexMale                      -0.05833    0.11586  -0.503
## SexMale:`Needle gauge`           -0.20319    0.48571  -0.418
## SexMale:PractitionerResident      0.40589    0.96106   0.422
## Age:`Needle gauge`                0.01671    0.06584   0.254
```
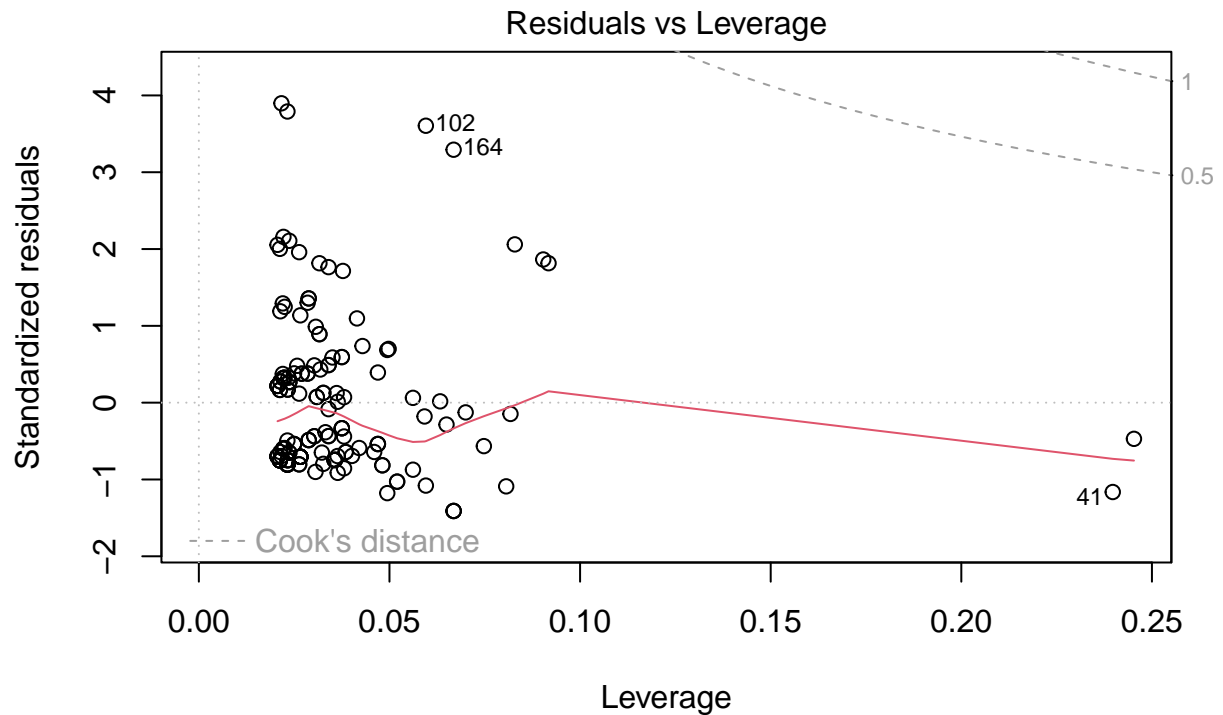
```
## Age:PractitionerResident                      -0.41277   0.14123  -2.923
## PractitionerResident:`Number of IV attempts` -0.56818   1.16120  -0.489
##                                               Pr(>|t|)
## (Intercept)                                     0.9343
## Age                                             0.8122
## SexMale                                         0.6588
## PractitionerResident                            0.0223 *
## `Number of IV attempts`                         0.0779 .
## `Needle gauge`                                  0.9332
## Age:SexMale                                     0.6154
## SexMale:`Needle gauge`                          0.6763
## SexMale:PractitionerResident                    0.6734
## Age:`Needle gauge`                              0.8000
## Age:PractitionerResident                        0.0040 **
## PractitionerResident:`Number of IV attempts`    0.6253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.155 on 152 degrees of freedom
## Multiple R-squared:  0.1474, Adjusted R-squared:  0.08571
## F-statistic: 2.389 on 11 and 152 DF,  p-value: 0.009291
```

```
plot(lm_model_without_interactions)
```
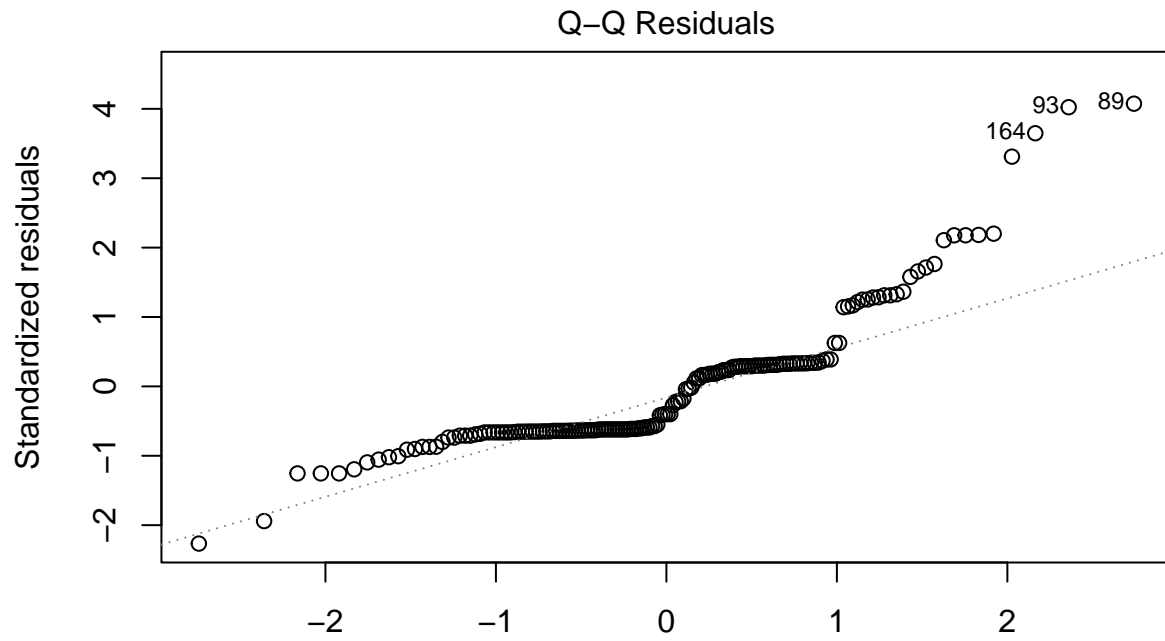
Print linear model without interactions

Residuals vs Fitted

lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Scale–Location

$\sqrt{|\text{Standardized residuals}|}$

Fitted values
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Residuals vs Leverage

Leverage
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

```
plot(lm_model_with_interactions)
```

**Print linear model with interactions**
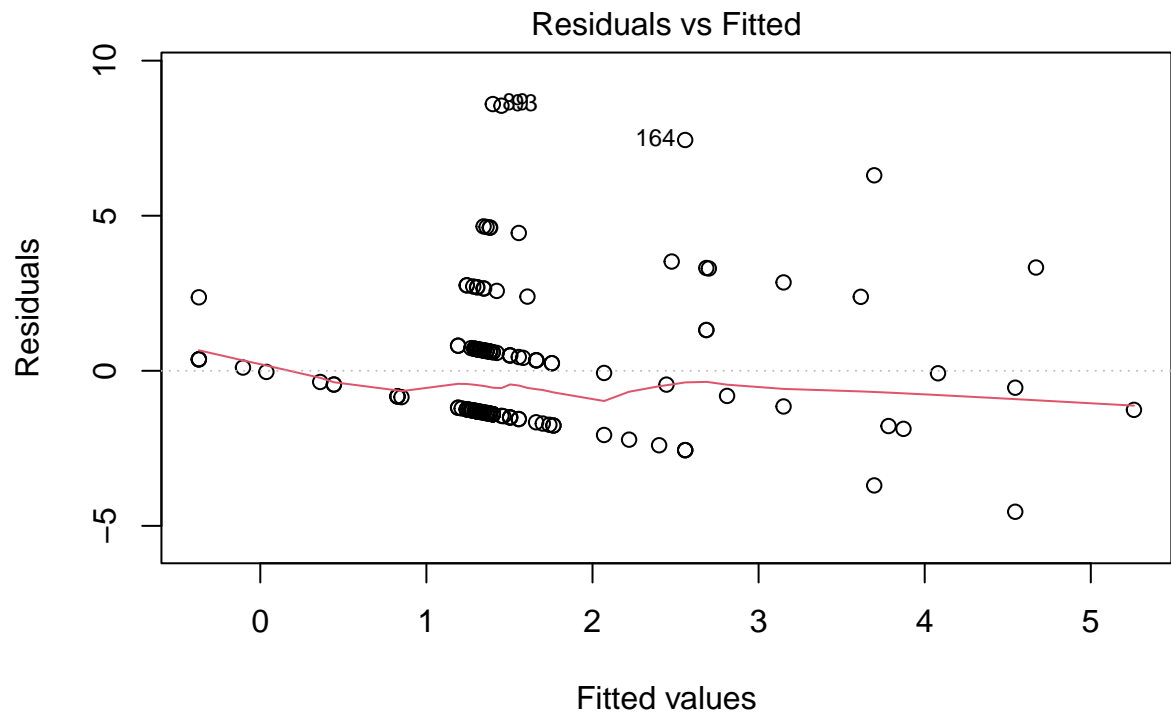
Residuals vs Fitted

Fitted values
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Scale–Location

√|Standardized residuals|

Fitted values
lm('FPS–R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...

Residuals vs Leverage
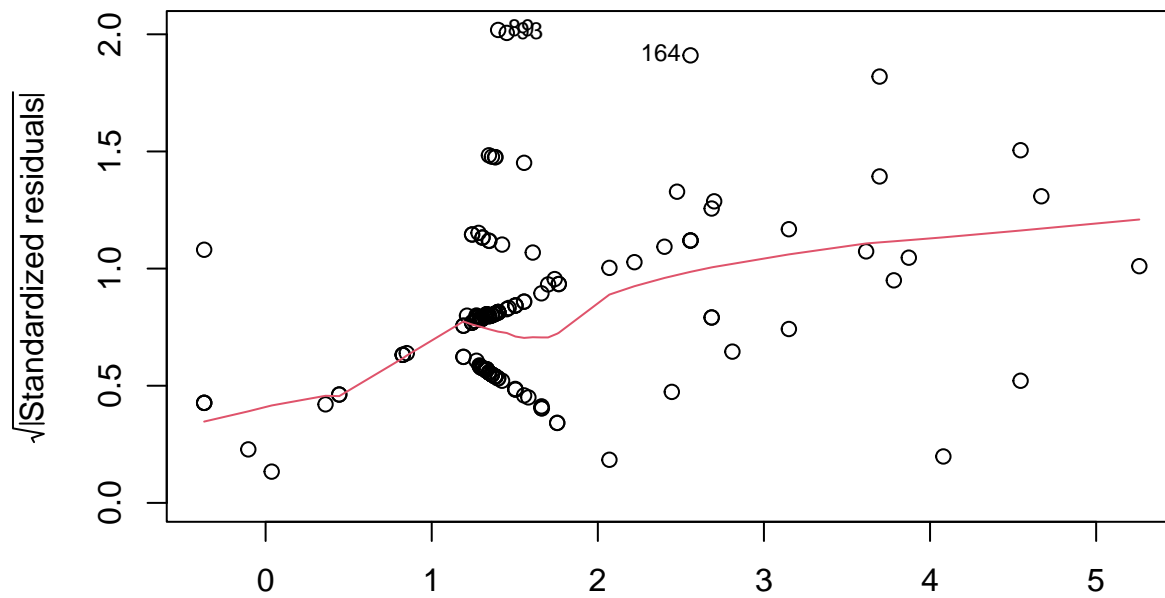
lm('FPS−R score' ~ Age + Sex + Practitioner + 'Number of IV attempts' + 'Ne ...
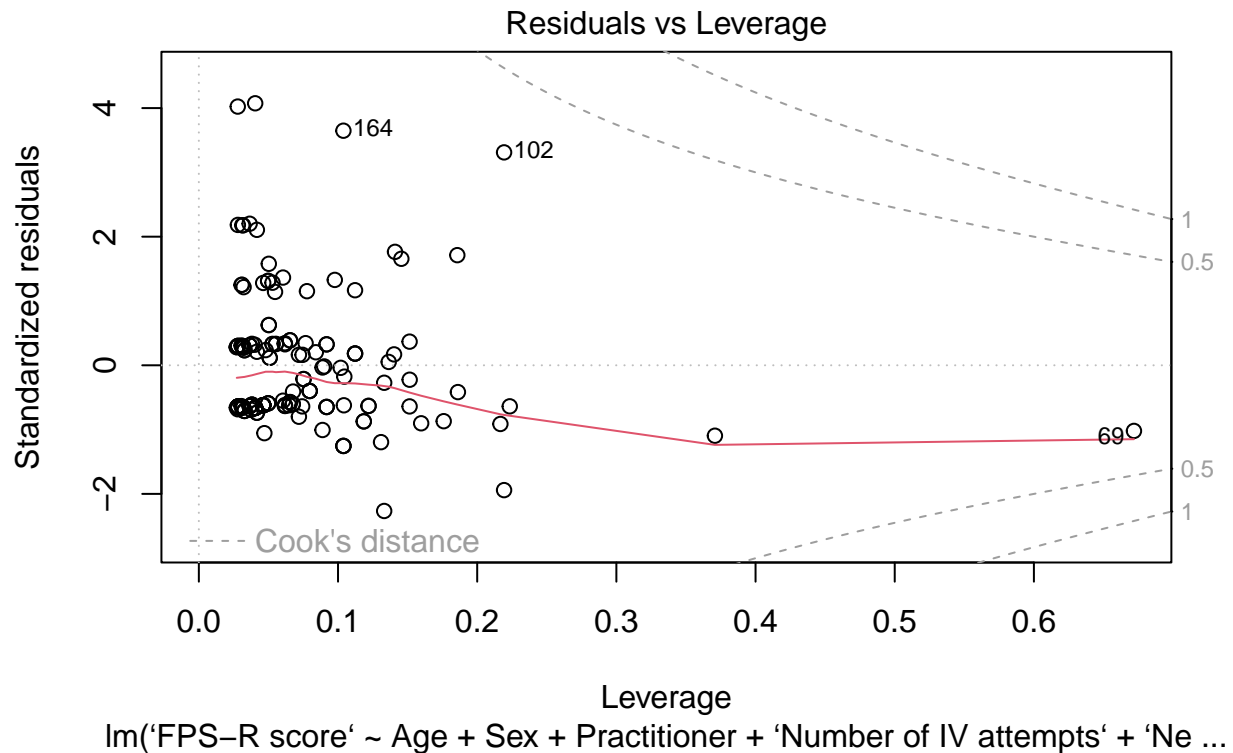
Objective 3

```r
# Pie chart draw function with percentages
draw_pie_chart <- function(all_categories, title) {
  # Count the occurrences of each category
  category_counts <- table(all_categories)

  # Calculate the percentage for each category
  percentages <- round(category_counts / sum(category_counts) * 100, 1)

  # Create labels that include both the category name and the percentage
  labels <- paste(names(category_counts), " - ", percentages, "%", sep="")

  # Create a pie chart with percentages
  pie(category_counts,
      main = title,
      col = rainbow(length(category_counts)),
      labels = labels,
      cex = 0.6)
}
```

```r
# Summarize "Any side effects observed?" column
side_effects_observed_summary <- data.frame(
  ifelse(original_data$`Any side effects observed?` == "No", "No Side Effect",
         ifelse(original_data$Randomization == 1, "Side Effect (Ametop alone)",
                "Side Effect (Ametop + spray)"))
)
```

```
# Create a pie chart for the observed side effects
draw_pie_chart(side_effects_observed_summary, "Observed Side Effects")
```

Pie chart for Observed side effects (No / Yes - Ametop alone / Yes - Ametop + spray)

Observed Side Effects
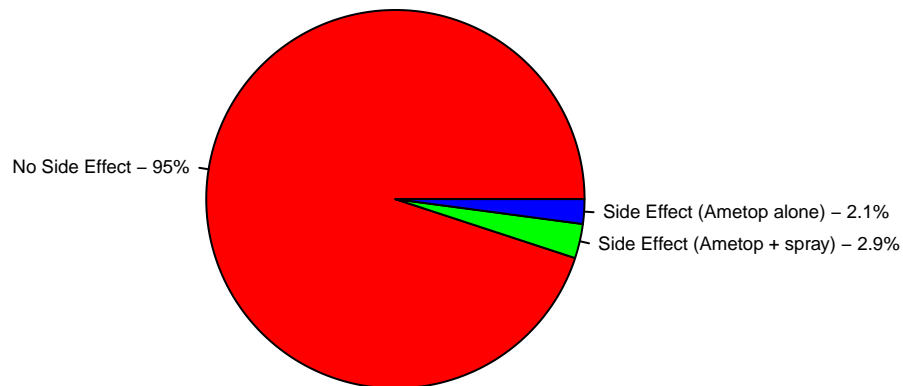

```
# Define the categorization function
categorize_effects <- function(effect) {
  effect_lower <- tolower(effect)
  categories <- list(
    Redness = "red|redness|pink",
    Itching = "itchy|pruritic|pruritis",
    Pain = "pain|painful",
    Puffiness = "puffy"
  )

  identified_categories <- character(0)

  # Negation patterns
  pattern_start <- "\\b(no|not)\\s+("
  pattern_middle <- ")\\b|(no|not)\\s+\\w+\\s+or\\s+("
  pattern_end <- ")"

  for (category_name in names(categories)) {
    symptom_pattern <- categories[[category_name]]

    # Check for the symptom but exclude if negated
    if (grepl(symptom_pattern, effect_lower)) {
      negation_pattern <- paste0(
        pattern_start, symptom_pattern, pattern_middle, symptom_pattern, pattern_end)
      if (!grepl(negation_pattern, effect_lower)) {
        identified_categories <- c(identified_categories, category_name)
      }
    }
  }

  # Return "Other" if no category is found
```

```r
  if (length(identified_categories) == 0) {
    return("Other")
  }

  return(identified_categories)
}
```

```r
# Create and print a list of categories for each side effect
categorize_and_print <- function(group_number) {
  side_effects_data <- data.frame(SideEffects = na.omit(original_data[
    original_data$Randomization == group_number,]$`Observed side effects`))
  side_effects_data$Categories <- sapply(side_effects_data$SideEffects,
                                          categorize_effects, USE.NAMES=FALSE)
  print(data.frame(Description = side_effects_data$SideEffects,
                   Categories = sapply(side_effects_data$Categories, paste, collapse = ", ")))
  return(side_effects_data)
}
```

```r
# Pie chart draw function for combined categories
draw_pie_chart_combined <- function(side_effects_data, title) {
  # Combination Categories
  all_categories <- sapply(side_effects_data$Categories, paste, collapse = " + ")

  # Use the draw_pie_chart function to create the pie chart
  draw_pie_chart(all_categories, title)
}
```

```r
# Pie chart draw function for individual categories
draw_pie_chart_individual <- function(side_effects_data, title) {
  # Flatten the list into a vector
  all_categories <- unlist(side_effects_data$Categories)

  # Use the draw_pie_chart function to create the pie chart
  draw_pie_chart(all_categories, title)
}
```

```r
# Get the side effects data for the control group
side_effects_data_control_group <- categorize_and_print(1)
```

Pie chart for observed side effects of Ametop alone (Redness, Itching, Pain, Puffiness)

```
##                                                   Description         Categories
## 1                          Slight redness of skin from ametop            Redness
## 2                             vey mild pink and pruritic skin.  Redness, Itching
## 3                                 pink skin and very mild itching            Redness
## 4 quite red & itchy skin from ametop. Underlying eczema.  Redness, Itching
## 5                          redness & raised area under ametop            Redness
```
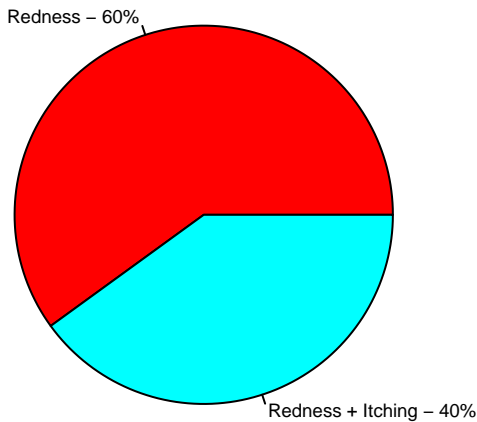
```r
# Pie Chart for combined categories
draw_pie_chart_combined(side_effects_data_control_group,
                        "Combination Side Effects Distribution - Ametop alone")
```
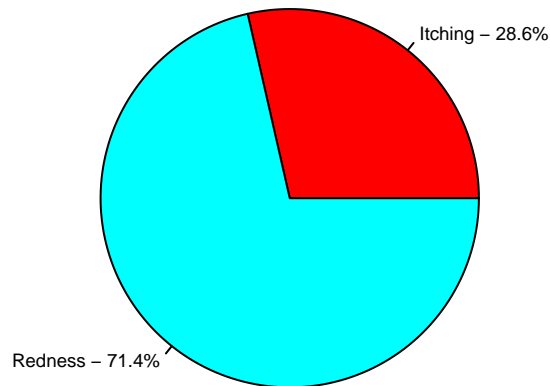
Combination Side Effects Distribution – Ametop alone

Redness – 60%

Redness + Itching – 40%

```
# Pie Chart for individual categories
draw_pie_chart_individual(side_effects_data_control_group,
                          "Individual Side Effects Distribution - Ametop alone")
```

Individual Side Effects Distribution – Ametop alone

Itching – 28.6%

Redness – 71.4%

```
# Get the side effects data for the study group
side_effects_data_study_group <- categorize_and_print(2)
```

Pie chart for observed side effects of Ametop + Spray (Redness, Itching, Pain, Puffiness)

```
##                                     Description         Categories
## 1          Some redness on skin, more than usual        Redness
## 2                  complained of pain on the spray.          Pain
## 3          slightly puffy hands after ametop removal    Puffiness
## 4                              redness from ametop        Redness
## 5  redness at ametop site. Not painful or pruritic.     Redness
## 6            quite itchy from ametop. Redness.  Redness, Itching
## 7        Slight redness from ametop. No pruritis.        Redness
```

```
# Pie Chart for combined categories
draw_pie_chart_combined(side_effects_data_study_group,
                        "Combination Side Effects Distribution – Ametop + Spray")
```

Combination Side Effects Distribution – Ametop + Spray
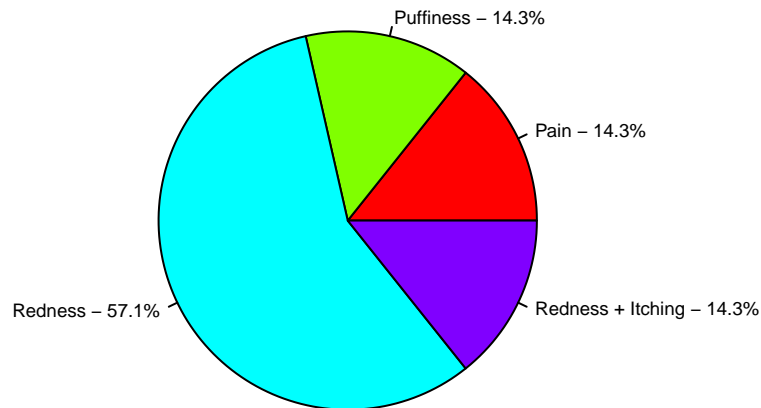

```
# Pie Chart for individual categories
draw_pie_chart_individual(side_effects_data_study_group,
                          "Individual Side Effects Distribution – Ametop + Spray")
```

Individual Side Effects Distribution – Ametop + Spray