In [1]:
```
library(gridExtra)
library(broom)
library(tidyverse)
library(car)
```

```
Warning message:
“package ‘lubridate’ was built under R version 4.4.2”
── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.2
── Conflicts ──────────────────────────────────────────── tidyverse_confl
icts() ──
✖ dplyr::combine() masks gridExtra::combine()
✖ dplyr::filter()  masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
Loading required package: carData


Attaching package: ‘car’


The following object is masked from ‘package:dplyr’:

    recode


The following object is masked from ‘package:purrr’:

    some
```

In [2]:
```
git_URL <- "https://raw.github.students.cs.ubc.ca/mfouzan/STAT301/main
raw_sleep_data <- read.csv(git_URL, sep = ",")

(raw_sleep_data)
```

A

Person.ID	Gender	Age	Occupation	Sleep.Duration	Quality.of.Sleep	Physica
<int>	<chr>	<int>	<chr>	<dbl>	<int>	
1	Male	27	Software Engineer	6.1	6	
2	Male	28	Doctor	6.2	6	

3	Male	28	Doctor	6.2	6
4	Male	28	Sales Representative	5.9	4
5	Male	28	Sales Representative	5.9	4
6	Male	28	Software Engineer	5.9	4
7	Male	29	Teacher	6.3	6
8	Male	29	Doctor	7.8	7
9	Male	29	Doctor	7.8	7
10	Male	29	Doctor	7.8	7
11	Male	29	Doctor	6.1	6
12	Male	29	Doctor	7.8	7
13	Male	29	Doctor	6.1	6
14	Male	29	Doctor	6.0	6
15	Male	29	Doctor	6.0	6
16	Male	29	Doctor	6.0	6
17	Female	29	Nurse	6.5	5
18	Male	29	Doctor	6.0	6
19	Female	29	Nurse	6.5	5
20	Male	30	Doctor	7.6	7
21	Male	30	Doctor	7.7	7
22	Male	30	Doctor	7.7	7
23	Male	30	Doctor	7.7	7
24	Male	30	Doctor	7.7	7
25	Male	30	Doctor	7.8	7
26	Male	30	Doctor	7.9	7
27	Male	30	Doctor	7.8	7
28	Male	30	Doctor	7.9	7
29	Male	30	Doctor	7.9	7
30	Male	30	Doctor	7.9	7

⋮	⋮	⋮	⋮	⋮	⋮
345	Female	57	Nurse	8.2	9
346	Female	57	Nurse	8.2	9
347	Female	57	Nurse	8.2	9
348	Female	57	Nurse	8.2	9
349	Female	57	Nurse	8.2	9
350	Female	57	Nurse	8.1	9
351	Female	57	Nurse	8.1	9
352	Female	57	Nurse	8.1	9
353	Female	58	Nurse	8.0	9
354	Female	58	Nurse	8.0	9
355	Female	58	Nurse	8.0	9
356	Female	58	Nurse	8.0	9
357	Female	58	Nurse	8.0	9
358	Female	58	Nurse	8.0	9
359	Female	59	Nurse	8.0	9
360	Female	59	Nurse	8.1	9
361	Female	59	Nurse	8.2	9
362	Female	59	Nurse	8.2	9
363	Female	59	Nurse	8.2	9
364	Female	59	Nurse	8.2	9
365	Female	59	Nurse	8.0	9
366	Female	59	Nurse	8.0	9
367	Female	59	Nurse	8.1	9
368	Female	59	Nurse	8.0	9
369	Female	59	Nurse	8.1	9
370	Female	59	Nurse	8.1	9
371	Female	59	Nurse	8.0	9
372	Female	59	Nurse	8.1	9
373	Female	59	Nurse	8.1	9
374	Female	59	Nurse	8.1	9

In [3]:
```
# No. of Rows; No. of Observations

rows = nrow(raw_sleep_data)
print(rows)
```

```
[1] 374
```

In [4]:
```
# No. of Columns; No. of Variables

col = ncol(raw_sleep_data)
print(col)
```

```
[1] 13
```

1. Data Description

The dataset contains **374 observations** and **13 variables** related to sleep health, lifestyle, and physiological metrics. Variables include demographic information (e.g., `Gender`, `Age`, `Occupation`), sleep metrics (e.g., `Sleep Duration`, `Quality of Sleep`), health indicators (e.g., `BMI Category`, `Blood Pressure`), and lifestyle factors (e.g., `Physical Activity Level`, `Daily Steps`). The data collection method is unspecified, but it appears to be synthetic or aggregated from self-reported surveys or health records.

Variable Summary Table

Variable Name	Type	Description
Person ID	Numeric	Unique identifier for individuals
Gender	Categorical	Male or Female
Age	Numeric	Age in years
Occupation	Categorical	Job role (e.g., Doctor, Engineer)
Sleep Duration	Numeric	Hours of sleep per night
Quality of Sleep	Numeric	Self-rated sleep quality (scale likely 1-10)
Physical Activity Level	Numeric	Daily activity minutes
Stress Level	Numeric	Stress score (scale likely 1-10)
BMI Category	Categorical	Obesity classification (e.g., Normal, Overweight)
Blood Pressure	String	Blood pressure measurement (e.g., "126/83")
Heart Rate	Numeric	Resting heart rate (bpm)
Daily Steps	Numeric	Number of steps per day

Sleep Disorder	Categorical	Sleep disorder diagnosis (e.g., None, Insomnia)

Data Source: Attached file `Sleep_health_and_lifestyle_dataset.csv`.
Citation: Tharmalingam, L. (2023) Sleep health and lifestyle dataset, Kaggle. Available at: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data?select=Sleep_health_and_lifestyle_dataset.csv

2: Predictive Research Question

a) Research Question

"How do stress level, physical activity level, occupation, and BMI category contribute to the variability in sleep duration(In Hours Per Night), as measured by a linear regression model?"

- **Response Variable**: `Sleep Duration` (numeric, hours).
- **Explanatory Variables**:
  - `Stress Level` (numeric: 1-10 scale),
  - `Physical Activity Level` (numeric: daily activity minutes),
  - `Occupation` (categorical: e.g., Doctor, Engineer),
  - `BMI Category` (categorical: Normal, Overweight, Obese).

---

b) Focus: Prediction

Goal: Build a regression model to predict sleep duration as a continuous numerical outcome.

Why Linear Regression Works Here:

- `Sleep Duration` is a **continuous numerical variable** (hours), making linear regression appropriate.
- The model will estimate the relationship between predictors and sleep duration, allowing for straightforward interpretation of coefficients.
- Alternative models (e.g., decision trees, gradient boosting, Exponential/log models) can also be explored if non-linear relationships exist.

3: Exploratory Data Analysis and Visualization

In [5]:
```
# Load and clean the dataset
sleep_data <- read.csv(git_URL) |>
  # Trim whitespace in categorical variables
  mutate(across(c(Gender, Occupation, BMI.Category, Sleep.Disorder), ~
  # Convert categorical variables to factors
```

```
  mutate(
    Gender = factor(Gender),
    Occupation = factor(Occupation),
    BMI.Category = factor(BMI.Category),
    Sleep.Disorder = factor(Sleep.Disorder)
  ) |>
  # Convert daily steps to thousands for cleaner interpretation
  mutate(Daily.Steps = Daily.Steps / 1000) |>
  # Remove person ID column
  select(-Person.ID) |>
  select(-Occupation)|>
  # Split blood pressure into numeric variables
  separate(Blood.Pressure, into = c("Systolic", "Diastolic"), sep = "/

# Verify structure
glimpse(sleep_data)
head(sleep_data)
tail(sleep_data)
```

```
Rows: 374
Columns: 12
$ Gender                  <fct> Male, Male, Male, Male, Male, Male, Mal
e, Male…
$ Age                     <int> 27, 28, 28, 28, 28, 28, 29, 29, 29, 29,
29, 29…
$ Sleep.Duration          <dbl> 6.1, 6.2, 6.2, 5.9, 5.9, 5.9, 6.3, 7.8,
7.8, 7…
$ Quality.of.Sleep        <int> 6, 6, 6, 4, 4, 4, 6, 7, 7, 7, 6, 7, 6,
6, 6, 6…
$ Physical.Activity.Level <int> 42, 60, 60, 30, 30, 30, 40, 75, 75, 75,
30, 75…
$ Stress.Level            <int> 6, 8, 8, 8, 8, 8, 7, 6, 6, 6, 8, 6, 8,
8, 8, 8…
$ BMI.Category            <fct> Overweight, Normal, Normal, Obese, Obes
e, Obes…
$ Systolic                <int> 126, 125, 125, 140, 140, 140, 140, 120,
120, 1…
$ Diastolic               <int> 83, 80, 80, 90, 90, 90, 90, 80, 80, 80,
80, 80…
$ Heart.Rate              <int> 77, 75, 75, 85, 85, 85, 82, 70, 70, 70,
70, 70…
$ Daily.Steps             <dbl> 4.2, 10.0, 10.0, 3.0, 3.0, 3.0, 3.5, 8.
0, 8.0,…
$ Sleep.Disorder          <fct> None, None, None, Sleep Apnea, Sleep Ap
nea, In…
```

A data.fram

	Gender	Age	Sleep.Duration	Quality.of.Sleep	Physical.Activity.Level	Stress.
	<fct>	<int>	<dbl>	<int>	<int>	
1	Male	27	6.1	6	42	
2	Male	28	6.2	6	60	
3	Male	28	6.2	6	60	
4	Male	28	5.9	4	30	
5	Male	28	5.9	4	30	
6	Male	28	5.9	4	30	

A data.fra

	Gender	Age	Sleep.Duration	Quality.of.Sleep	Physical.Activity.Level	Stre
	<fct>	<int>	<dbl>	<int>	<int>	
369	Female	59	8.1	9	75	
370	Female	59	8.1	9	75	
371	Female	59	8.0	9	75	
372	Female	59	8.1	9	75	
373	Female	59	8.1	9	75	
374	Female	59	8.1	9	75	

In [6]:

```
# Printing summary of the cleaned data

summary(sleep_data)
```

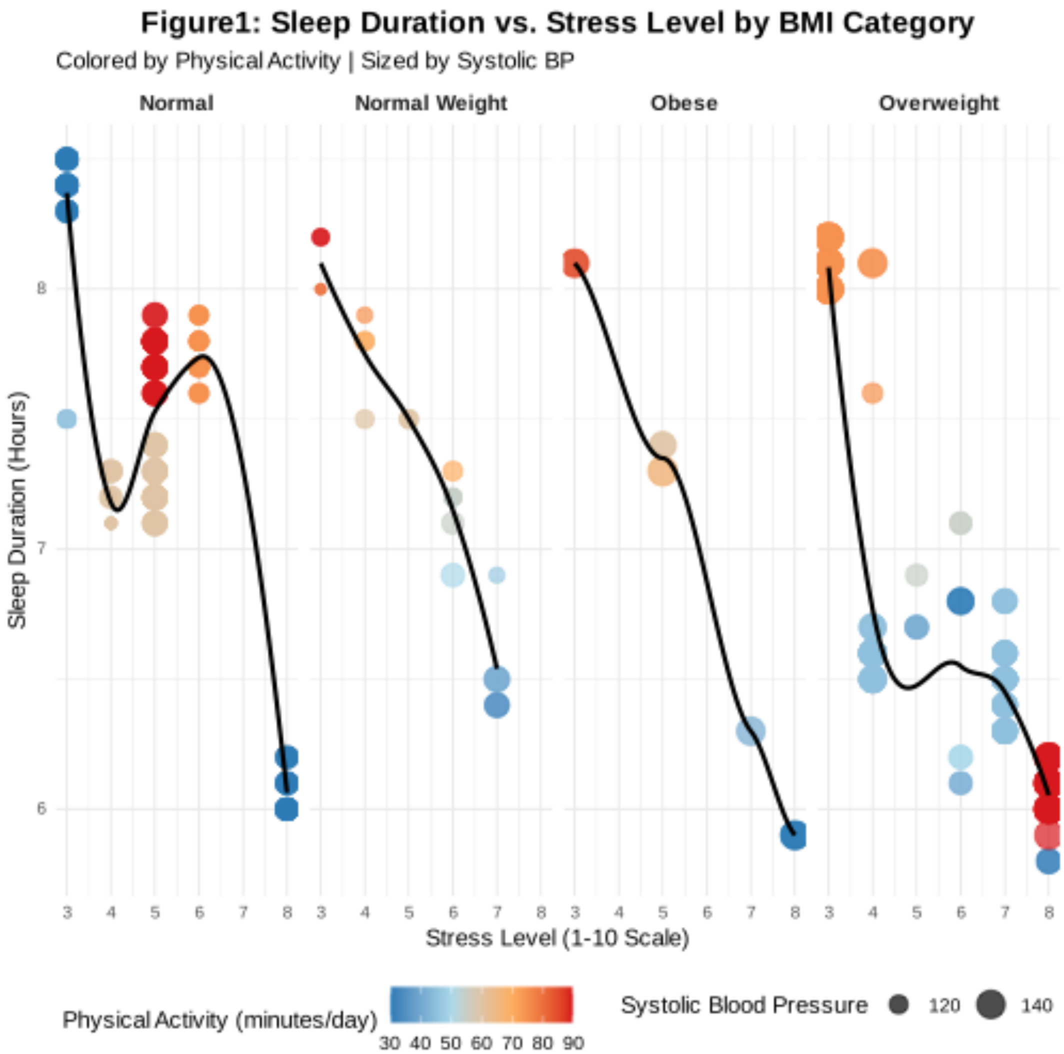
```
    Gender          Age        Sleep.Duration  Quality.of.Sleep
 Female:185   Min.   :27.00   Min.   :5.800   Min.   :4.000  
 Male  :189   1st Qu.:35.25   1st Qu.:6.400   1st Qu.:6.000  
              Median :43.00   Median :7.200   Median :7.000  
              Mean   :42.18   Mean   :7.132   Mean   :7.313  
              3rd Qu.:50.00   3rd Qu.:7.800   3rd Qu.:8.000  
              Max.   :59.00   Max.   :8.500   Max.   :9.000  
 Physical.Activity.Level  Stress.Level          BMI.Category    Systoli
c
 Min.   :30.00           Min.   :3.000   Normal       :195   Min.   :11
5.0
 1st Qu.:45.00           1st Qu.:4.000   Normal Weight: 21   1st Qu.:12
5.0
 Median :60.00           Median :5.000   Obese        : 10   Median :13
0.0
 Mean   :59.17           Mean   :5.385   Overweight   :148   Mean   :12
8.6
 3rd Qu.:75.00           3rd Qu.:7.000                       3rd Qu.:13
5.0
 Max.   :90.00           Max.   :8.000                       Max.   :14
2.0
   Diastolic       Heart.Rate     Daily.Steps          Sleep.Disorder
 Min.   :75.00   Min.   :65.00   Min.   : 3.000   Insomnia    : 77  
 1st Qu.:80.00   1st Qu.:68.00   1st Qu.: 5.600   None        :219  
 Median :85.00   Median :70.00   Median : 7.000   Sleep Apnea: 78  
 Mean   :84.65   Mean   :70.17   Mean   : 6.817                     
 3rd Qu.:90.00   3rd Qu.:72.00   3rd Qu.: 8.000                     
 Max.   :95.00   Max.   :86.00   Max.   :10.000                     
```

In [7]:

```
# Custom theme for readability
theme_custom <- theme_minimal() +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    strip.text = element_text(face = "bold", size = 10),
    legend.position = "bottom",
    axis.text = element_text(size = 8)
  )

# Create the plot
sleep_data |>
  ggplot(aes(x = Stress.Level, y = Sleep.Duration)) +
  # Scatterplot: Color = Physical Activity, Size = Systolic BP
  geom_point(aes(color = Physical.Activity.Level, size = Systolic), al
  # Trendlines for each BMI group
  geom_smooth(aes(group = BMI.Category), method = "loess", color = "bl
  # Facet by BMI Category
  facet_wrap(~ BMI.Category, nrow = 1) +
  # Color gradient for physical activity
  scale_color_gradientn(
    name = "Physical Activity (minutes/day)",
    colors = c("#2c7bb6", "#abd9e9", "#fdae61", "#d7191c")
  ) +
  # Size scale for systolic blood pressure
```

```
  scale_size_continuous(
    name = "Systolic Blood Pressure",
    range = c(2, 6),
    breaks = c(120, 140, 160)
  ) +
  # Labels and titles
  labs(
    title = "Figure1: Sleep Duration vs. Stress Level by BMI Category"
    subtitle = "Colored by Physical Activity | Sized by Systolic BP",
    x = "Stress Level (1–10 Scale)",
    y = "Sleep Duration (Hours)"
  ) +
  theme_custom
```

```
`geom_smooth()` using formula = 'y ~ x'
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“pseudoinverse used at 6”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“neighborhood radius 2”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“reciprocal condition number  5.4192e-17”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“pseudoinverse used at 2.975”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“neighborhood radius 4.025”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“reciprocal condition number  9.5238e-17”
Warning message in simpleLoess(y, x, w, span, degree = degree, parametr
ic = parametric, :
“There are other near singularities as well. 9.1506”
```



Key Interpretation

Figure 1. Relationship between stress level and sleep duration, stratified by BMI category. Points are colored by daily physical activity level (minutes) and sized by resting heart rate (bpm). Three key observations emerge:

1. **BMI Moderation**:
   - The negative relationship between stress and sleep duration is strongest in the **Obese** group (steepest slope).
   - Normal BMI individuals maintain relatively stable sleep duration across stress levels.
2. **Protective Effect of Activity**:
   - Higher physical activity (yellow points) correlates with better sleep retention at high stress levels, particularly in the Overweight group.
3. **Physiological Warning Signs**:

- Larger points (elevated heart rates) cluster in high-stress/short-sleep regions, suggesting a triad of:
  - Chronic stress
  - Sleep deprivation
  - Cardiovascular strain

Implications for Modeling:

- Interaction terms between stress and BMI category should be explored.
- Heart rate and physical activity show non-linear relationships worthy of spline terms.

Methods and Plan (Assignment 2)

Proposed Method:
We will employ **multiple linear regression** to investigate the relationship between sleep duration (response variable) and various explanatory variables. Following initial model fitting, we will perform **model selection** using Variance Inflation Factor (VIF) analysis and **backward elimination** to identify the most relevant predictors. Residual diagnostics will assess assumptions and model validity.

---

1. Why is this method appropriate?

- Sleep duration is a continuous variable, making regression suitable for quantifying associations.
- Multiple regression allows simultaneous evaluation of multiple predictors (e.g., physical activity, screen time, stress levels) while controlling for confounding effects.
- Model selection (VIF/backward elimination) optimizes model simplicity and predictive accuracy by removing redundant or non-significant variables.

2. Key Assumptions:

- **Linearity:** Relationships between predictors and response are linear.
- **Independence:** Observations are independent (e.g., data not repeated measures).
- **Homoscedasticity:** Residual variance is constant across predicted values.
- **Normality:** Residuals are approximately normally distributed.
- **No Multicollinearity:** Predictors are not highly correlated (addressed via VIF).

3. Limitations and Weaknesses:

- **Linearity Assumption:** Non-linear relationships may be missed; transformations (e.g., log) could mitigate this.
- **Causality:** Regression identifies associations, not causation.
- **Outliers/Influential Points:** Extreme values may disproportionately affect results.
- **Data Quality:** Results depend on accurate measurement of variables (e.g., self-reported data may introduce bias).

Next Steps:

1. Fit an initial full model with all predictors.
2. Use VIF to detect multicollinearity and remove variables with VIF > 5.
3. Apply backward elimination (p-value threshold: 0.05) to refine the model.
4. Validate assumptions via residual plots (e.g., residuals vs. fitted values, Q-Q plots).
5. Interpret final model coefficients to answer the research question.

Code Implementation

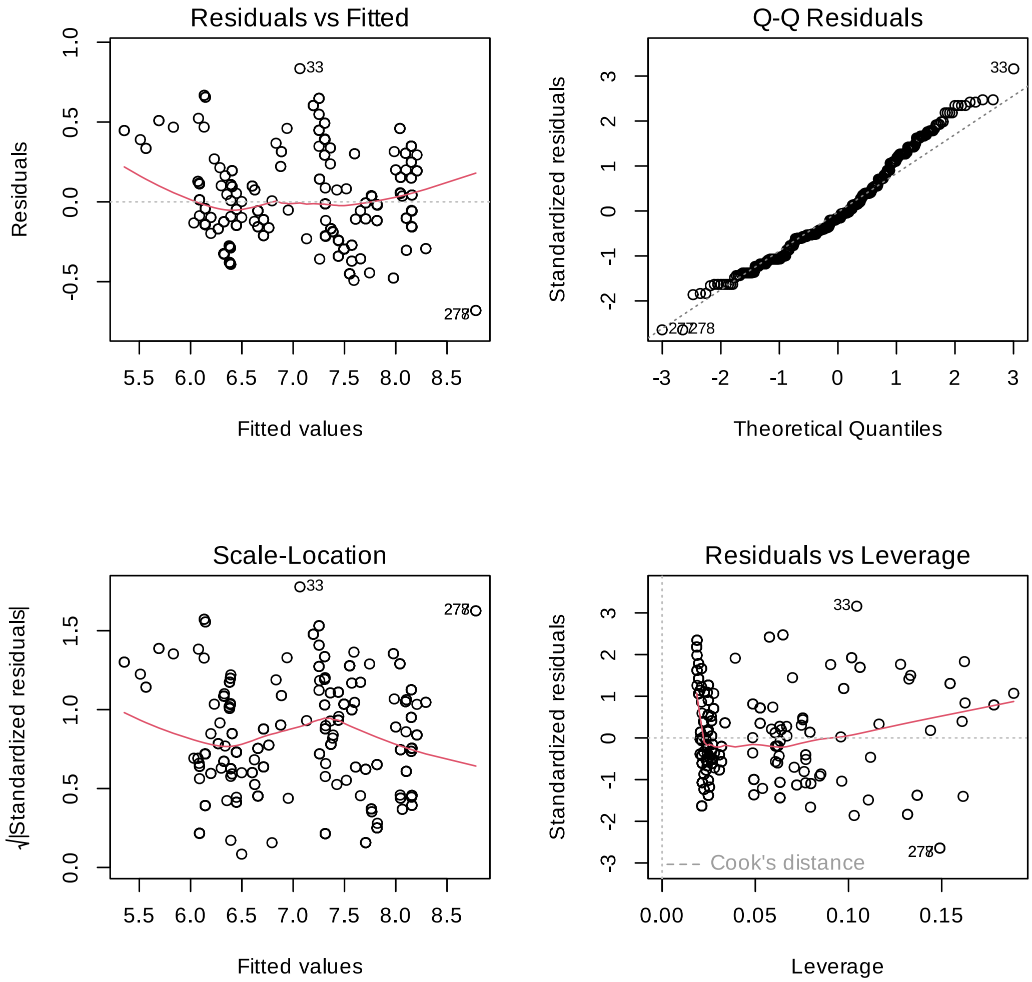
In [8]:

```
# Fitting full model with sleep duration as response and all explanato
full_model <- lm(Sleep.Duration ~ ., data = sleep_data)

# Print initial model summary
cat("Initial Full Model Summary:\n")
print(summary(full_model))
```

```
Initial Full Model Summary:

Call:
lm(formula = Sleep.Duration ~ ., data = sleep_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.68042 -0.16786 -0.04472  0.15315  0.83454 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                8.060878   0.924323   8.721  < 2e-16 ***
GenderMale                 0.538781   0.049525  10.879  < 2e-16 ***
Age                        0.055239   0.005138  10.750  < 2e-16 ***
Quality.of.Sleep           0.054415   0.053286   1.021 0.307849    
Physical.Activity.Level    0.005855   0.001673   3.501 0.000522 ***
Stress.Level              -0.245344   0.032093  -7.645 1.93e-13 ***
BMI.CategoryNormal Weight -0.131488   0.070874  -1.855 0.064381 .  
BMI.CategoryObese         -0.645782   0.172924  -3.734 0.000219 ***
BMI.CategoryOverweight    -0.831307   0.084152  -9.879  < 2e-16 ***
Systolic                  -0.136503   0.015041  -9.076  < 2e-16 ***
Diastolic                  0.164952   0.020335   8.112 8.00e-15 ***
Heart.Rate                 0.026406   0.009658   2.734 0.006564 ** 
Daily.Steps               -0.133519   0.022569  -5.916 7.70e-09 ***
Sleep.DisorderNone         0.050956   0.058319   0.874 0.382838    
Sleep.DisorderSleep Apnea  0.054779   0.064760   0.846 0.398184    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.279 on 359 degrees of freedom
Multiple R-squared:  0.8816,	Adjusted R-squared:  0.877 
F-statistic:   191 on 14 and 359 DF,  p-value: < 2.2e-16
```

In [9]:

```
# Check residual plots
par(mfrow = c(2, 2))
plot(full_model)
```



The residuls dont seem to have a random pattern, but in fact there is a little obvious pattern that concerns multicolleniarity

In [10]:
```
# Check for multicollinearity
vif_values <- vif(full_model)
print(vif_values)
```

```
                           GVIF Df GVIF^(1/(2*Df))
Gender                 2.944807  1        1.716044
Age                    9.514085  1        3.084491
Quality.of.Sleep      19.486443  1        4.414345
Physical.Activity.Level  5.814697  1        2.411368
Stress.Level          15.535683  1        3.941533
BMI.Category          27.106903  3        1.733192
Systolic              65.054063  1        8.065610
Diastolic             75.202061  1        8.671912
Heart.Rate             7.642231  1        2.764459
Daily.Steps            6.386890  1        2.527230
Sleep.Disorder         6.906444  2        1.621114
```

Remove variables that have large VIF Values.

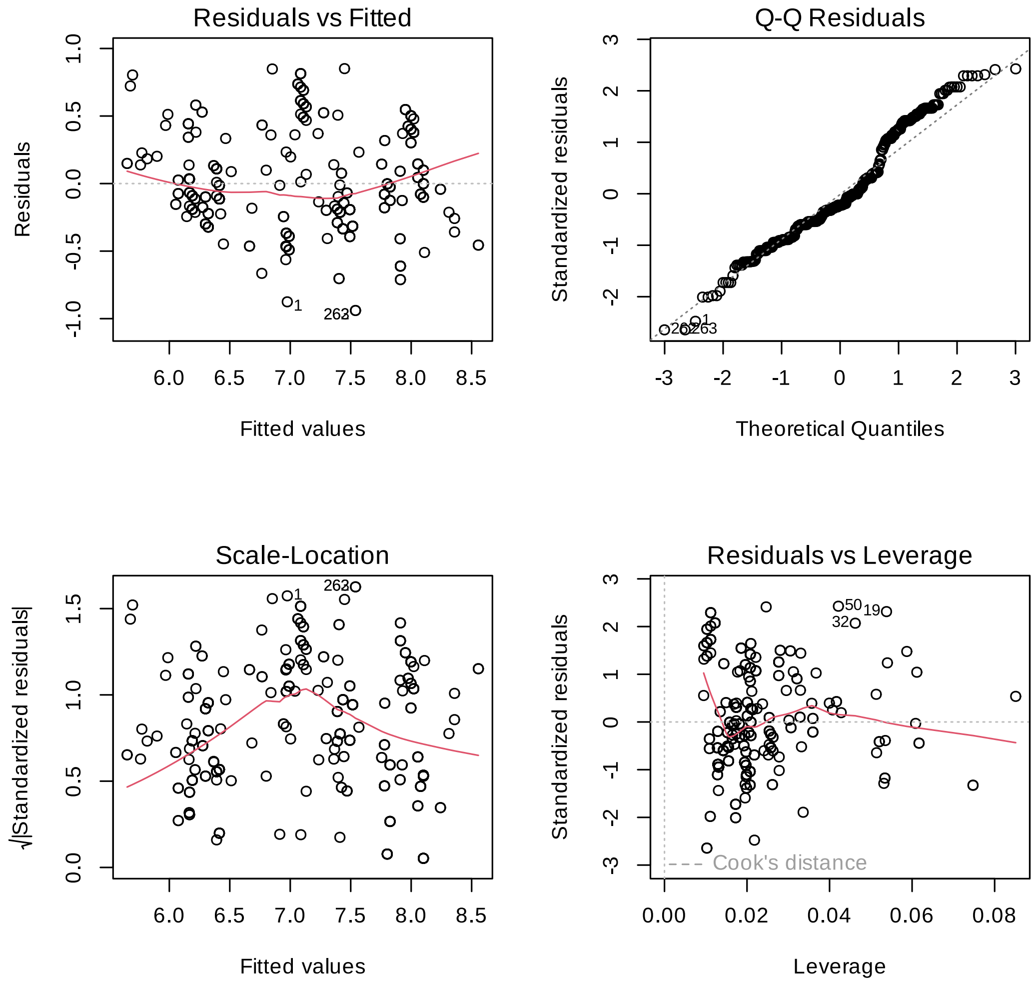
In [11]:
```
# Reduced data, removing the variables
reduced_data <- sleep_data %>%
  select(-Systolic, -Diastolic, -Heart.Rate, -BMI.Category, -Quality.o
```

In [12]:
```
# Fitting the reduced model
reduced_model <- lm(Sleep.Duration ~ ., data = reduced_data)

print(vif(reduced_model))
```

```
                            GVIF Df GVIF^(1/(2*Df))
Gender                  1.949089  1        1.396098
Age                     2.250244  1        1.500081
Physical.Activity.Level 3.486064  1        1.867100
Stress.Level            2.089803  1        1.445615
Daily.Steps             3.235986  1        1.798885
Sleep.Disorder          2.443517  2        1.250270
```

In [13]:
```
# Backward elimination
final_model <- step(reduced_model, direction = "backward", trace = 0)
```

In [14]:
```
#  Plotting the final model
par(mfrow = c(2, 2))
plot(final_model)
```



In [15]:
```
# Summary of the final model
summary(final_model)
```

```
Call:
lm(formula = Sleep.Duration ~ Gender + Age + Physical.Activity.Level + 
    Stress.Level + Daily.Steps + Sleep.Disorder, data = reduced_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.93959 -0.21241 -0.07799  0.20063  0.85099 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                7.266641   0.198547  36.599  < 2e-16 ***
GenderMale                 0.440653   0.051579   8.543 3.57e-16 ***
Age                        0.022998   0.003199   7.189 3.70e-12 ***
Physical.Activity.Level    0.007249   0.001658   4.372 1.60e-05 ***
Stress.Level              -0.331366   0.015068 -21.991  < 2e-16 ***
Daily.Steps               -0.051977   0.020565  -2.527   0.0119 *  
Sleep.DisorderNone         0.549178   0.054366  10.102  < 2e-16 ***
Sleep.DisorderSleep Apnea  0.291748   0.068412   4.265 2.55e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3572 on 366 degrees of freedom
Multiple R-squared:  0.8022,	Adjusted R-squared:  0.7984 
F-statistic: 212.1 on 7 and 366 DF,  p-value: < 2.2e-16
```

Interpretation of the Results

This is the summary of a multiple linear regression model predicting **Sleep Duration** based on various independent variables. Here's the interpretation of key components:

1. Model Formula:

$$ \text{Sleep.Duration} \sim \text{Gender} + \text{Age} + \text{Physical.Activity.Level} + \text{Stress.Level} + \text{Daily.Steps} + \text{Sleep.Disorder} $$

This means that **Sleep Duration** is being predicted based on gender, age, physical activity level, stress level, daily steps, and sleep disorder type.

---

2. Coefficients Interpretation:

Each coefficient represents the expected change in **Sleep Duration (in hours)** for a one-unit increase in the predictor, holding all else constant.

Predictor	Estimate	Interpretation

(Intercept)	7.266	The baseline sleep duration when all predictors are at their reference values.
Gender (Male)	0.441	Males sleep **0.441 hours (≈26 minutes) more** than females, on average.
Age	0.023	Each additional year of age increases sleep duration by **0.023 hours (≈1.4 minutes)**.
Physical Activity Level	0.007	Higher physical activity level is associated with **0.007 hours (≈25 seconds) more sleep**.
Stress Level	-0.331	For each unit increase in stress, sleep duration decreases by **0.331 hours (≈20 minutes)**.
Daily Steps	-0.052	More steps per day are linked to a **slight decrease** in sleep duration.
Sleep Disorder (None vs. Disorder)	0.549	People **without sleep disorders** sleep **0.549 hours (≈33 minutes) more** than those with sleep disorders.
Sleep Apnea	0.292	People with **sleep apnea** sleep **0.292 hours (≈17 minutes) more** than those with other disorders.

---

3. Statistical Significance:

- **Stars (`***` , ` `,` *`) indicate significance levels**:
  - `***` (p < 0.001) → Highly significant
  - `**` (p < 0.01) → Significant
  - `*` (p < 0.05) → Moderately significant
- **All variables except Daily Steps (p = 0.0119) are highly significant (p < 0.001), meaning they strongly impact Sleep Duration**.

---

4. Model Fit:

- **Multiple R^2 = 0.8022** → The model explains **80.22% of the variance** in Sleep Duration.
- **Adjusted R^2 = 0.7984** → Adjusted for the number of predictors, still strong.
- **F-statistic = 212.1, p-value < 2.2e-16** → The overall model is statistically significant.

```
In [ ]:
```