Reproducible Report Template

Project 1

Student/s 520254163 520179176 510022956 510590259

University of Sydney | Math1005 | Lab Class 03 G11

1 Executive Summary

The purpose of this report is to explore whether the different parts of the human body can influence each other by analyze the data of height, weight, gender and fingers based the collected dataset. Graphical and numerical summaries are available to answer the three research questions "Is there a relationship between Height and Fingerprint Area ?", "Is there a relationship between Height and Weight ?" and "the relationship between gender and fingerprint temperature" The results of this study found out that There is a moderate correlation between height and finger area as well as height and weight, strongly supported by linear regression. Finally, we also observed that male tend to have warmer fingers than female, as indicated by our comparative boxplot graphs. These relevant results allow stakeholders and organizations to get a better understanding of the degree of mutual influence between body parts and apply the conclusion in life.

2 Full Report

2.1 Initial Data Analysis (IDA)

2.1.1 IDA sourse

This data comes from Loughborough University's online data site (repository.lboro.ac.uk), author by Beth McMurchie, George Torrens, Paul Kelly, posted on 09.01.2019.

All the following research question data are from this dataset and are strictly based on the original data, no changes will be made. The data comes from the university's own website, and the data set is relatively complete, and the data volume is large. Although there are missing data on age, these aging data will not be used in the report as the primary study object, so it is judged that the data is valid for the report.

The reliability of the data needs to be considered because the background of the data survey is not informed, and the source of the data is not clearly expressed. The three authors are all university lecturers, and the age of the test subjects is mostly concentrated between 18-25 years old. Based on these information, it can be known that the respondents are mainly students. This is just a speculation based on data and will not be applied to the actual report. Nonetheless, due to the uncertainty of the respondents and the possible idiosyncrasies, we cannot conclude that the data are of high social generality and can be considered as confounding variables explaining outliers. The data are overwhelmingly numeric, so the influence of native language on the subjects can be ruled out.

This data is related to physical ergonomics, which may be useful for medical research or for height prediction. Potential stakeholders include, hospitals, fingerprint scholars, growth assessment physicians, and people who interested in height issues.

2.1.2 variables

The classification the data variables are as shown:

```
## [1] "C:/Users/pc/Desktop"
```

```
## 'data.frame':    200 obs. of  11 variables:
##  $ Participant.Number                     : int  101 102 103 104 105 106 107 108 109 110 ...
##  $ Gender                                 : chr  "Male" "Male" "Male" "Male" ...
##  $ Age                                    : int  NA NA NA NA 18 20 NA NA NA NA ...
##  $ Dominant.Hand                          : chr  "Right" "Right" "Right" "Right" ...
##  $ Height..cm...average.of.3.measurments. : num  174 202 182 184 181 ...
##  $ Weight..kg...average.of.3.measurements.: num  70 99 82 75 80.3 59 71 81 74 85 ...
##  $ Finger.temperature                     : int  34 30 29 29 29 32 28 28 26 32 ...
##  $ Fingerprint.Height..mm.                : num  19.8 24 20 23.2 22.7 24.3 20 20.6 24.1 22.6
...
##  $ Fingerprint.Width..mm.                 : num  13.7 14.1 13.7 14 15 14 15 13.4 16.3 14.6
...
##  $ Fingerprint.Area..mm2.                 : num  241 279 224 282 287 ...
##  $ Fingerprint.Circumference..mm.         : num  57.7 62.7 55.5 63.3 62.7 65.6 58 58 67.6 63.
6 ...
```

There are 200 rows in the data, and each row represents one of the 200 respondents. There are 11 columns in the data, and each column represents a tested dependent variable.

The key variables used in this report are:

Gender: is a string. The gender of the respondents was divided into male and female.

Height..cm…average.of.3.measurments: is a numeral represents the respondent's height in cm. This data is taken as the average of three heights evaluated.

Weight..kg…average.of.3.measurements: is a numeral represents the respondent's weight in kg. This data is taken as the average of three weights evaluated.

Fingerprint.Area..mm2: is an integer represents the area of the respondent's finger in mm2.

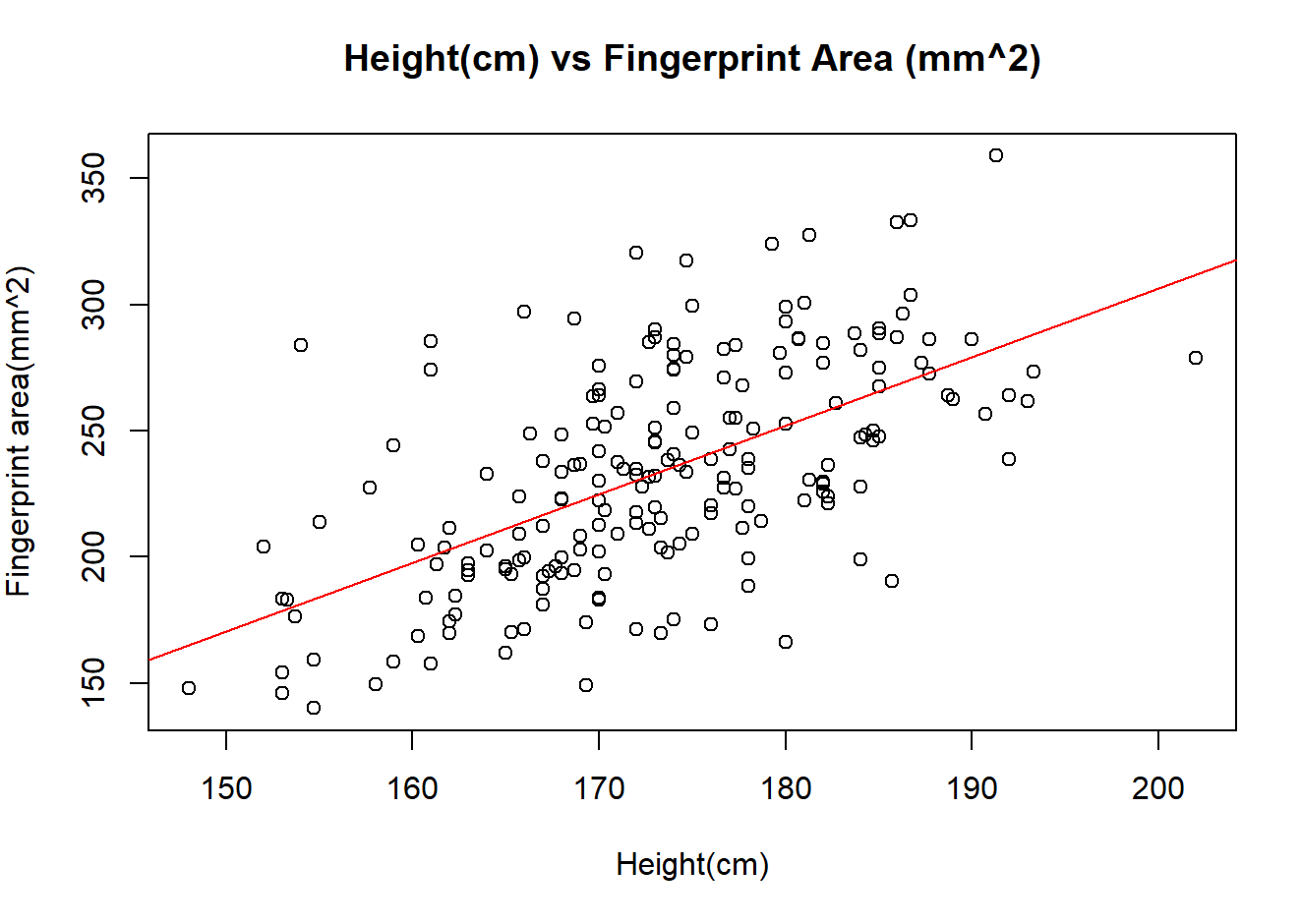
Fingertip.Temperature…C: is an integer represents the temperature of the respondent's finger in °C.

We decided not to use finger length and width in the report, because our chosen data - the area of finger is enough to represent those two data (length and width multiplied to equal area). Although the data for finger temperature is missing in several places, it is possible to ignore the missing parts of the column due to the considerable data size.

2.2 Research Question 1: In the research one, we would like to determine the relationship between Height and Fingerprint Area about participants

To begin, we will use scatter plots with regression line and correlation coefficients from people's height to fingerprint area to determine whether there is linear relationship between them.

2.2.1 Regression line




```
##
## Call:
## lm(formula = Area ~ Height)
##
## Coefficients:
## (Intercept)       Height
##     -237.50         2.72
```

```
## [1] 0.5934453
```

The equation for the linear regression line was y=-237.50+2.72x. The correlation coefficient was 0.59, which suggests that a linear model could be appropriate for the relationship between the height and fingerprint percentages. To investigate further into whether a linear model is truly appropriate, a residual plot can be created.

2.2.2 Residuals

```
##      0%     25%     50%     75%    100%
## 140.000 199.650 232.700 267.475 358.900
```

```
## [1] 67.825
```

```
##      25%
## 97.9125
```

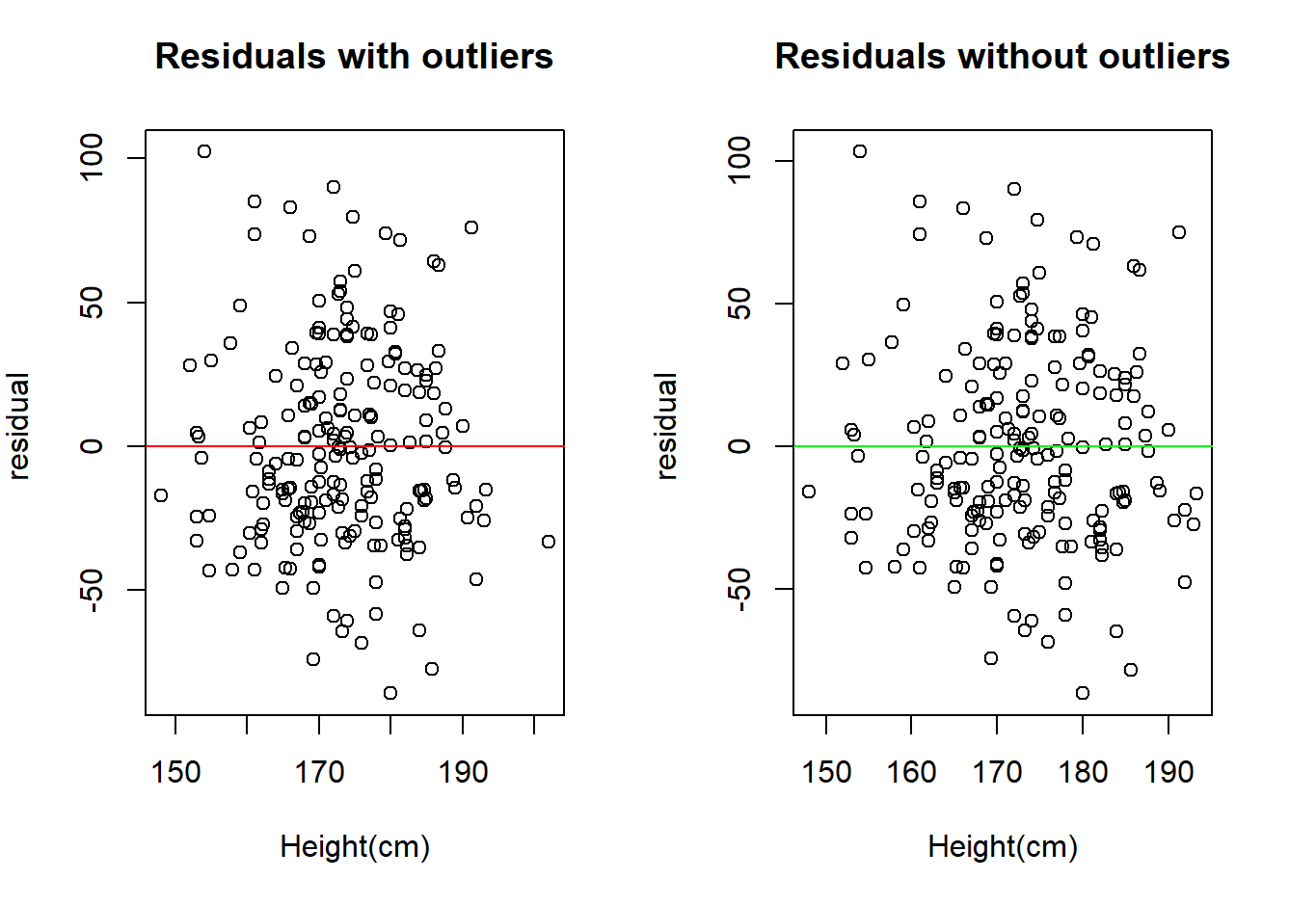
```
##      75%
## 369.2125
```

```
##      0%     25%     50%     75%    100%
## 148.000 167.225 173.000 180.000 202.000
```

```
## [1] 12.775
```

```
##      25%
## 148.0625
```

```
##      75%
## 199.1625
```



The plot diagram above depicts the relationship between height and residuals with(left side) or without outliers (right side). When the residual plots are compared, the x-axis scale changes as larger values are removed, making the cluster more central in the plot without outliers. The central cluster spreads out more with a narrower x-axis, making it easier to observe equal spread within vertical strips. In addition, there appears to be no discernible pattern in the plot, indicating homoscedasticity.

On the other hand, there aren't many outliers in this data; only a small percentage of the data above 199.1625 cm are outliers. By contrast, The residual plot without outliers appears more ideal because its spread and homoscedasticity are more obvious. In comparison to the plot with outliers, it strongly suggests that a linear model is appropriate and that the regression line can be used for predictions. This leads to the conclusion that there is a linear relationship between a person's height and fingerprint area, with the fingerprint area increasing as the height increases and vice versa.

To summarize, when determining whether two sets of data are linear, we must examine not only the scatter plot of the data and the value of the correlation coefficient, but also the residual plot diagram. In the first study, a linear regression model and residual plot were used to determine that there is a medium positive linear relationship between a person's fingerprint area and height in this data set. We can then conduct a more detailed investigation to see if there is a linear relationship between other physical indicators of a person.

2.3 Research Question 2: In the research2 we would like to find the relationship of Height(cm) and Weight(kg)

Height(cm) vs Weight(kg) The purpose of the second research is to analyze the data of participants' height and weight and determine whether and what kind of correlation they are.In this part, we plan to use the combination of scatter plot and regression line as well as residual plot to analyze and compare the data.

2.3.1 Regression line

```
## [1] 12.775
```

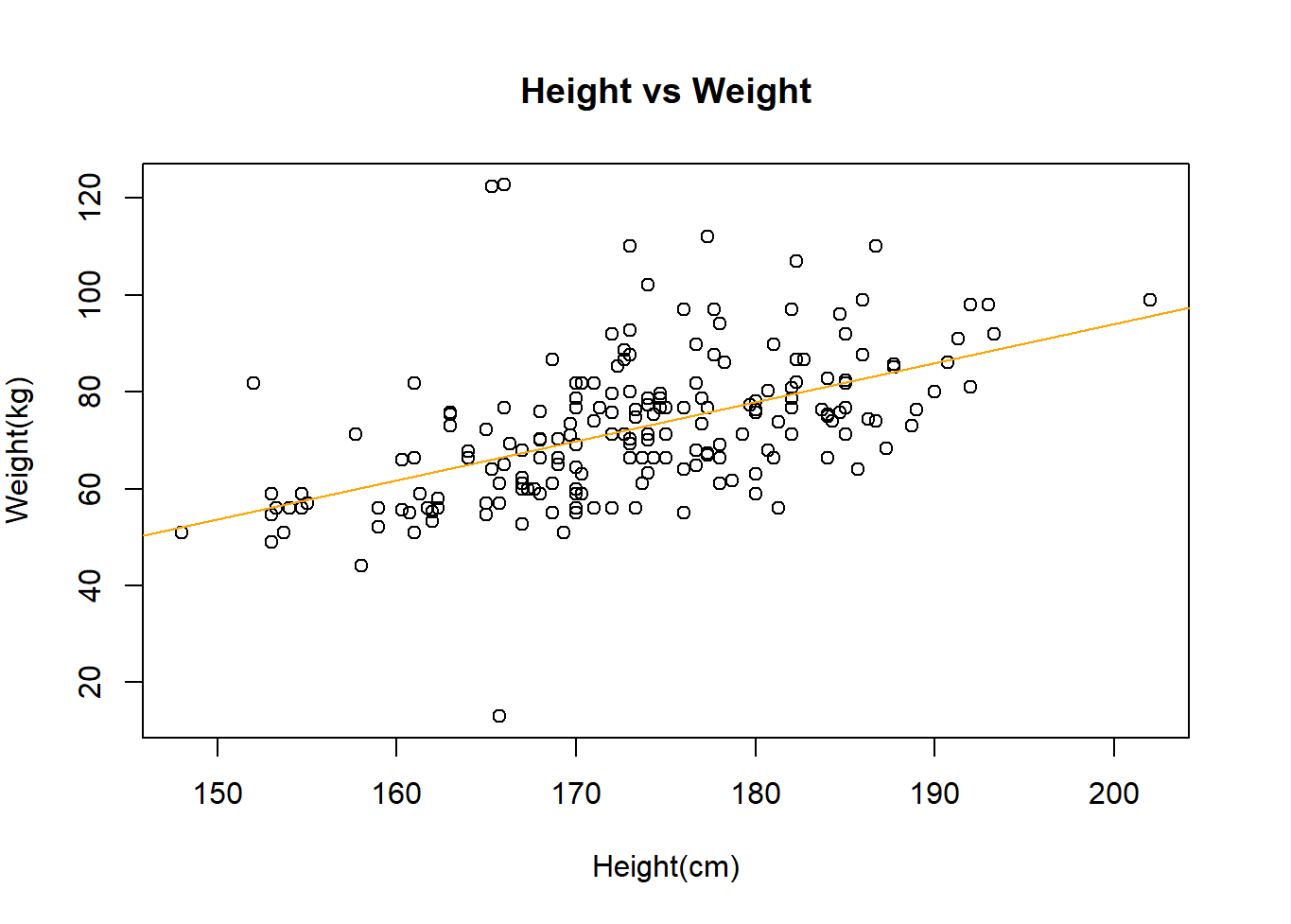
```
##      25%
## 148.0625
```

```
##      75%
## 199.1625
```

```
## [1] 19.075
```

```
##     25%
## 32.3875
```

```
##      75%
## 108.6875
```




```
##
## Call:
## lm(formula = w ~ h)
##
## Coefficients:
## (Intercept)            h
##     -67.183        0.806
```

```
## [1] 0.5156449
```

The linear regression equation is: y=0.806x-67.183 and the correlation coefficient was 0.5156649. This suggests that linear models may be applicable to the relationship between height and weight in participants. Residual plots can be created to further investigate whether the linear model is really appropriate.

2.3.2 Residual

```
## [1] 12.775
```

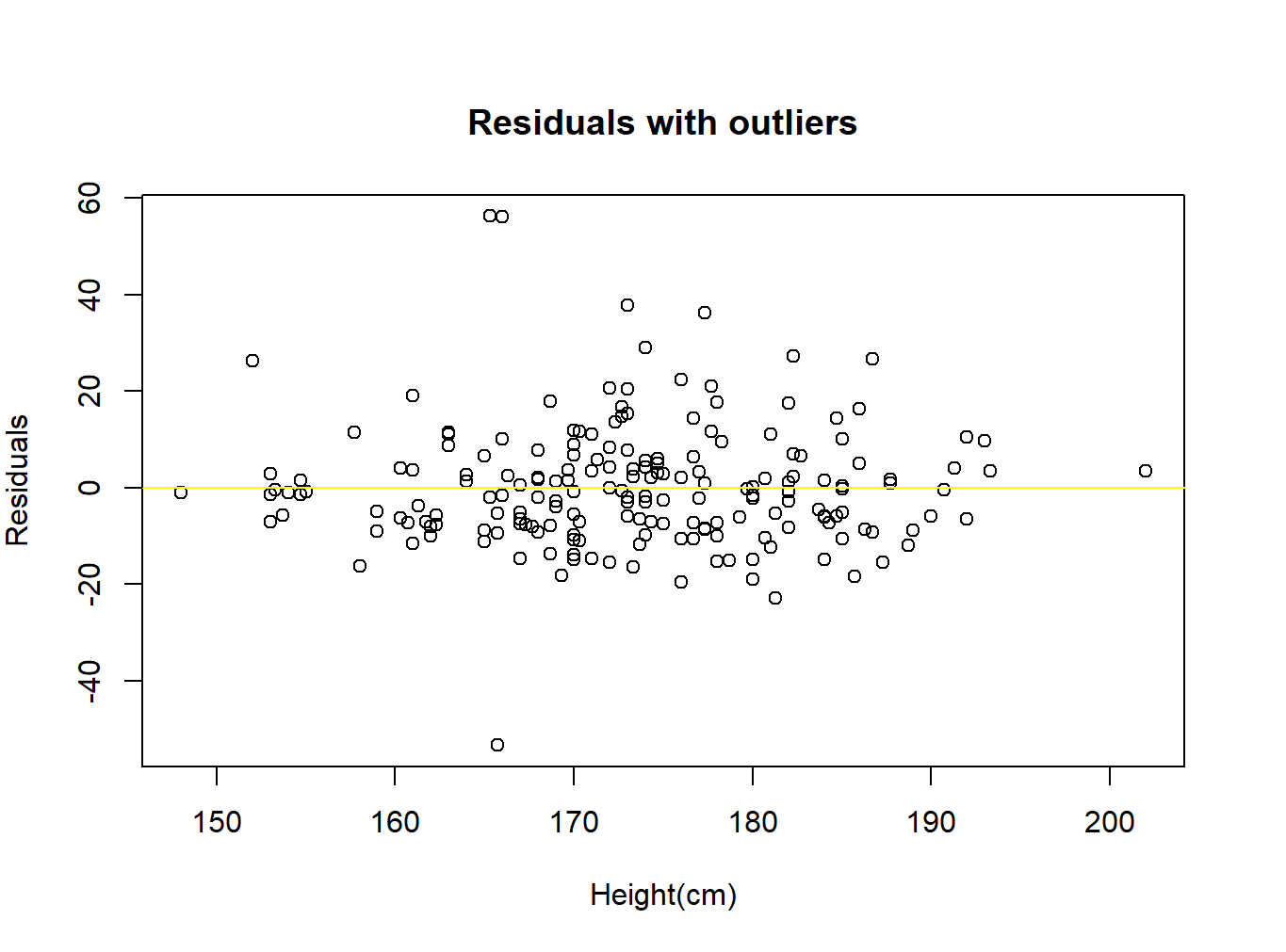
```
##      25%
## 148.0625
```

```
##      75%
## 199.1625
```

```
## [1] 19.075
```

```
##     25%
## 32.3875
```

```
##      75%
## 108.6875
```




```
##      0%     25%     50%     75%    100%
## 148.000 167.225 173.000 180.000 202.000
```

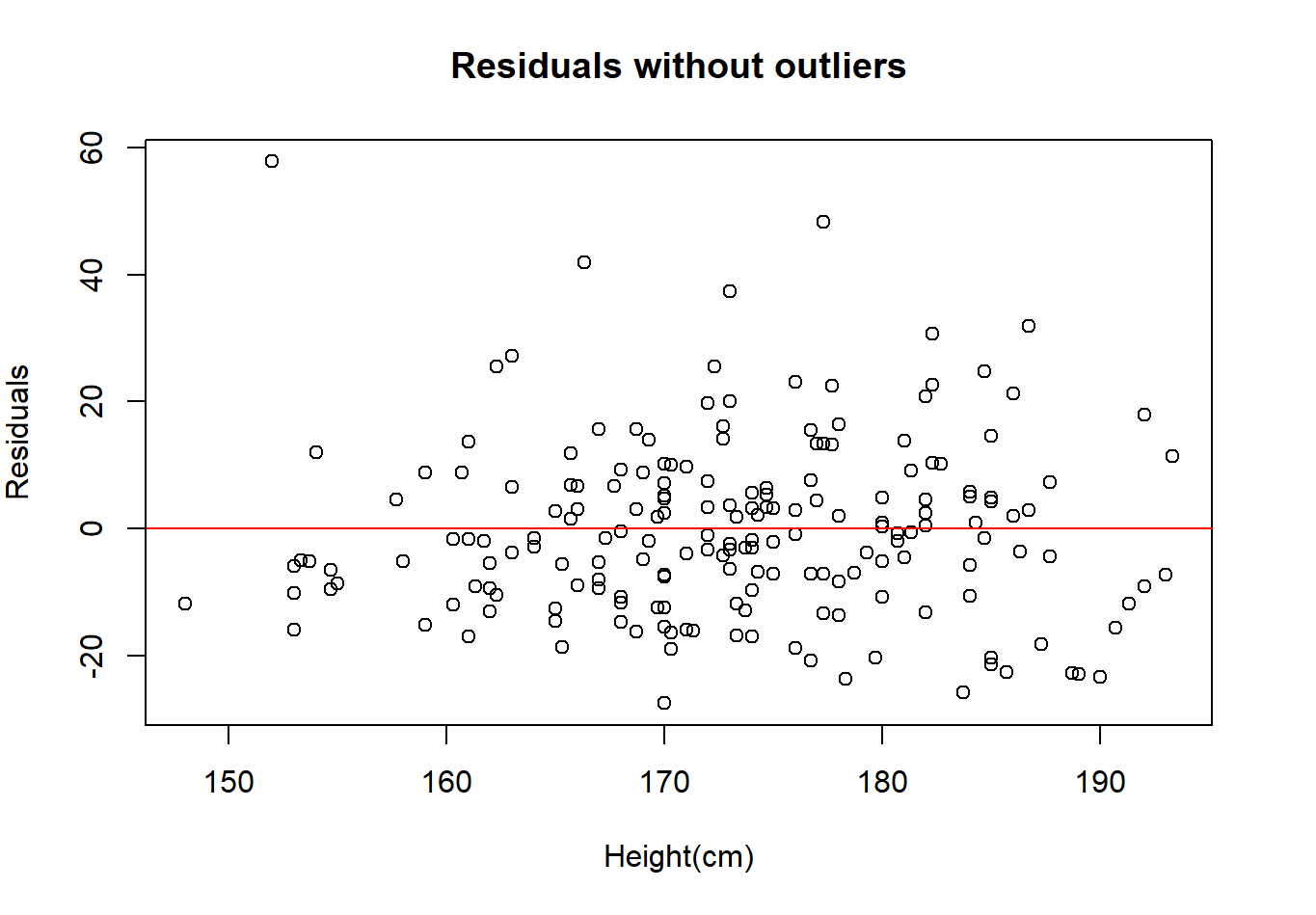
```
## [1] 12.775
```

```
##      25%
## 148.0625
```

```
##      0%     25%     50%     75%    100%
##  13.000  61.000  71.300  80.075 122.700
```

```
## [1] 19.075
```

```
##     25%
## 32.3875
```



The use of residual maps allows for a better comparison of X-axis scale changes. Looking at the cluster located between the boundary of y=-40 and y=40 shows that the narrower the X-axis the easier it is to observe the distribution within the vertical band. However, homoscedasticity is not clearly observed in the graph without removing outliers.

In contrast, the residual map without outliers is more ideal. It has clear homoscedastic property and spread. It is a stronger indication that the linear model is appropriate and that the regression line can be used for prediction than a graph with outliers. So we can conclude that there is a linear relationship between the height and weight of the participants. Weight increases with height and decreases vice versa. There is a positive correlation.

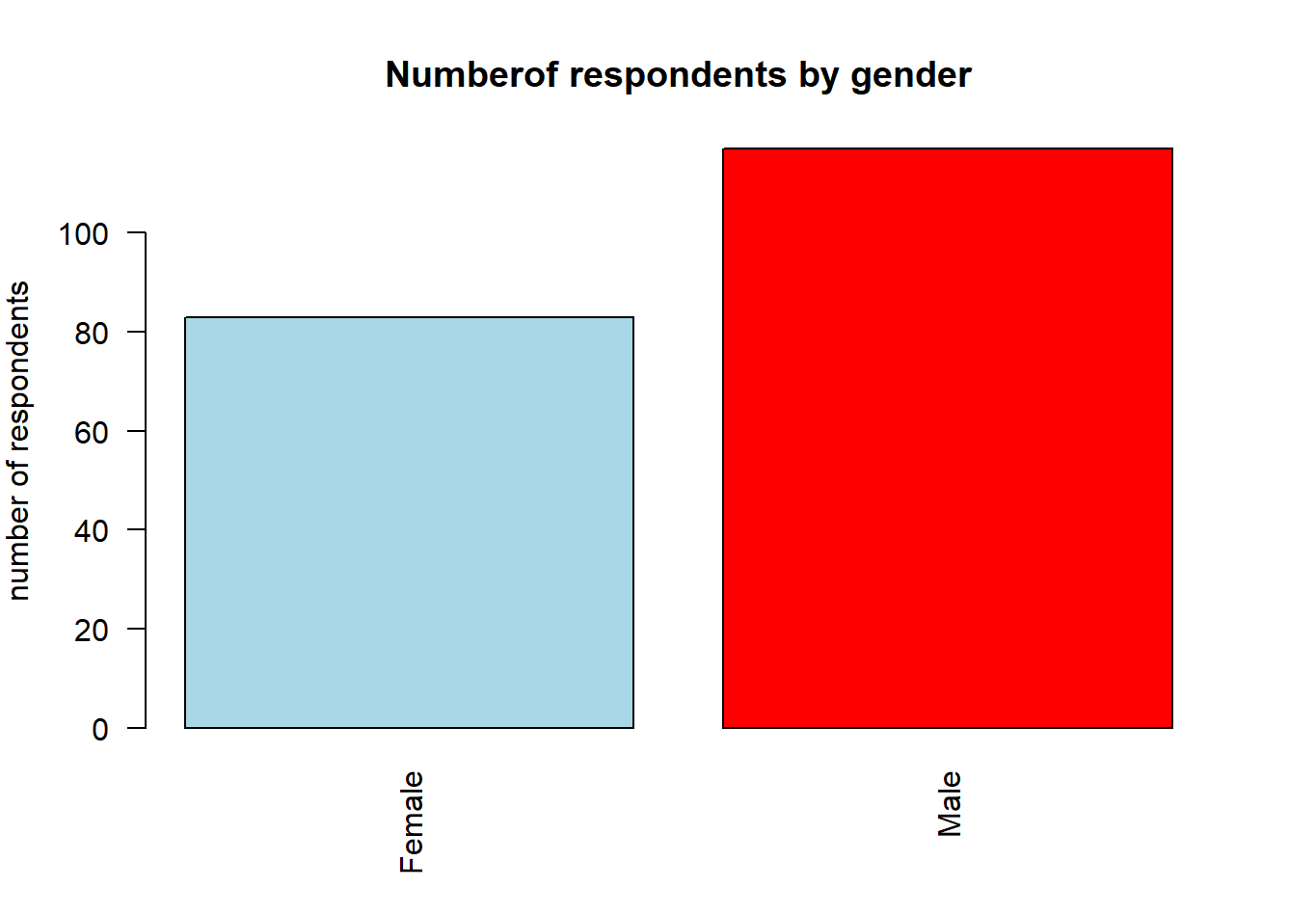
To explore the relationship between height and weight of participants, we can perform comparative analyses by using linear regression models and residual plots. However, it should be noted that NA needs to be removed before use, and when drawing residual graphs, attention should be paid to removing outliers and analyzing and comparing residual graphs with abnormal values. In conclusion, the above analysis can show that there is a positive linear relationship between the height and weight of participants in this data set. This conclusion has also been reached by other external studies that have highlighted the effect of height on weight in the wider population.

2.4 Further research:In further research, we want to determine the differences in fingerprint temperature between genders

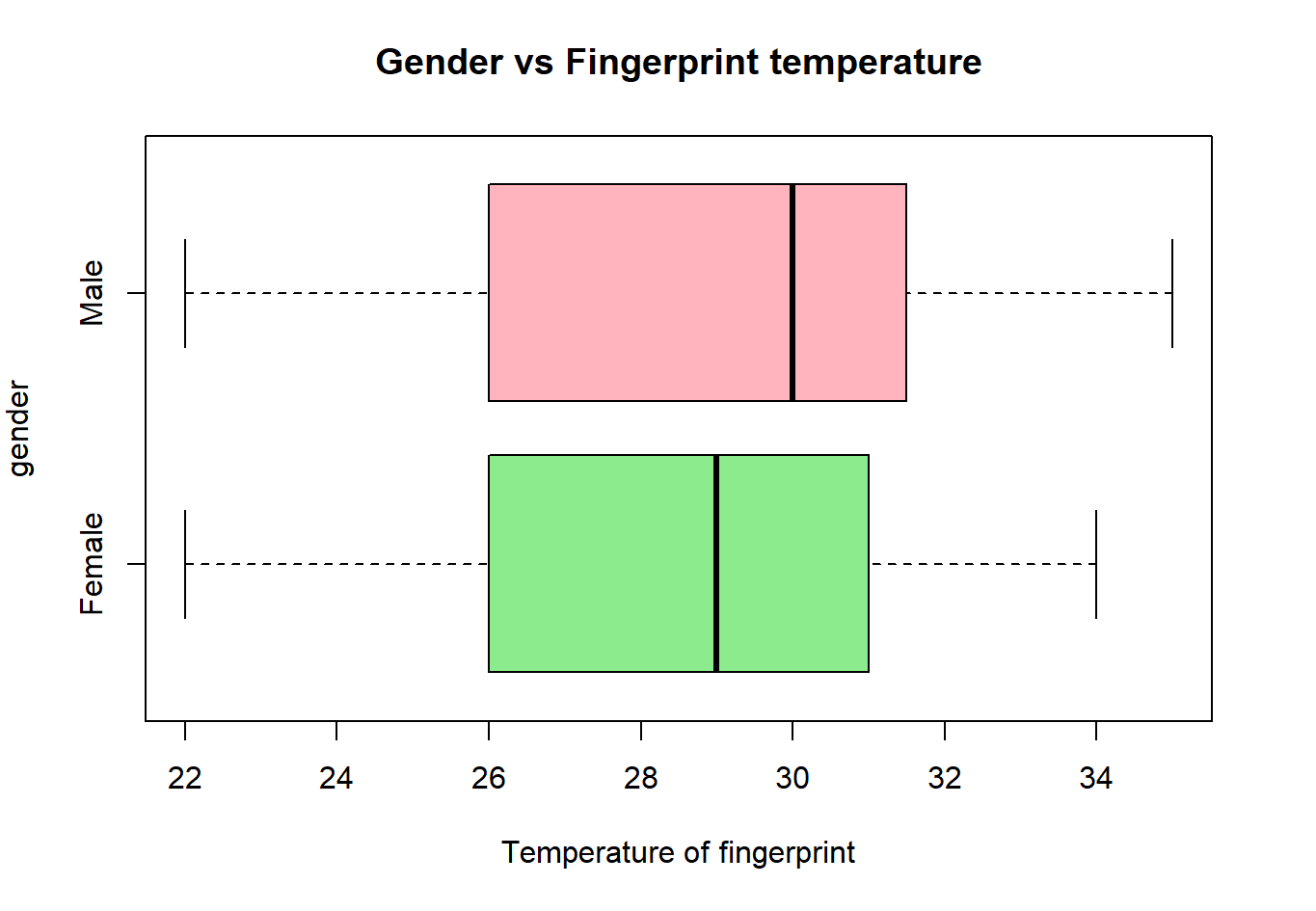
Gender vs Fingertip Temperature (°C)

```
## 'data.frame':    200 obs. of  11 variables:
##  $ Participant.Number                     : int  101 102 103 104 105 106 107 108 109 110 ...
##  $ Gender                                 : chr  "Male" "Male" "Male" "Male" ...
##  $ Age                                    : int  NA NA NA NA 18 20 NA NA NA NA ...
##  $ Dominant.Hand                          : chr  "Right" "Right" "Right" "Right" ...
##  $ Height..cm...average.of.3.measurments. : num  174 202 182 184 181 ...
##  $ Weight..kg...average.of.3.measurements.: num  70 99 82 75 80.3 59 71 81 74 85 ...
##  $ Finger.temperature                     : int  34 30 29 29 29 32 28 28 26 32 ...
##  $ Fingerprint.Height..mm.                : num  19.8 24 20 23.2 22.7 24.3 20 20.6 24.1 22.6
...
##  $ Fingerprint.Width..mm.                 : num  13.7 14.1 13.7 14 15 14 15 13.4 16.3 14.6
...
##  $ Fingerprint.Area..mm2.                 : num  241 279 224 282 287 ...
##  $ Fingerprint.Circumference..mm.         : num  57.7 62.7 55.5 63.3 62.7 65.6 58 58 67.6 63.
6 ...
```

```
## [1] "character"
```




```
## [1] "integer"
```



Initially,It can be clearly seen from Figure 1 that there are more male respondents than female respondents.Figure 2 shows that for both men and women, the lowest temperature for fingerprints is 25 degrees Celsius.In general, men's fingerprints are warmer than women's.The main reason is that men's body temperatures are higher than women's.

A report has shown that The body's metabolism is responsible for the production of energy, including heat. Although men and women maintain an internal body temperature of 98.6 degrees, men typically have more muscle mass and generate more heat by using more calories to fuel those extra muscles. When that heat evaporates, it warms up their skin, their clothes and the air just above the surface of their skin."Basically, men generate their own little heat islands, kind of like walking space heaters," this is what Dr. Danoff said. "But since women typically have less muscle mass and evaporate less heat through the pores in their skin, they might feel colder than men in a room with the same air temperature."Some studies have shown that while men feel comfortable in rooms with the thermostat set at 72 degrees, women tend to feel comfortable in rooms with the thermostat set at 77(Johnson,2022).

Objectively speaking, men undertake more physical and intellectual labor, so men's metabolism is more vigorous than women, high metabolism, can release more energy, body temperature is naturally higher than women, this is why men eat more than women. From the perspective of tissue structure, men not only have a larger proportion of muscle tissue than women, but also have more blood in the body than women, and the contraction of muscle tissue is the main means of raising body temperature. This is how people shake when they are exposed to cold (strong contractions of muscles produce a lot of heat to fight the cold). A higher blood ratio than a woman indicates a man. This is why men's fingerprints are generally warmer than women's

3 References

Johnson, B. (2022). Why Do Women Typically Feel Colder Than Men?. Find a DO | Doctors of Osteopathic Medicine. Retrieved 22 September 2022, from https://findado.osteopathic.org/why-do-women-typically-feel-colder-than-men.

McMurchie, Beth; Torrens, George; Kelly, Paul (2019): Height, weight and fingerprint measurements collected from 200 participants. Loughborough University. Dataset. https://doi.org/10.17028/rd.lboro.7539206.v1.