EDA_PS3_GROUP13

Ayesha Bhimdiwala, Pulkit Maloo and Vaishnavi

1/29/2018

```
library(ggplot2)
library(NHANES)
library(broom)
```

```
## Warning: package 'broom' was built under R version 3.4.3
```

INTRODUCTION

High systolic blood pressure is a strong predictor of heart attacks and strokes. We are trying to find a relationship of average systolic blood pressure with age, height, and weight respectively using the NHANES data which consists of the following variables:

BPSysAve (the average of three measurements of systolic blood pressure)

Age (in years; 80 or older is recorded as 80)

Weight (in kilograms)

Height (in centimeters)

Gender (male or female)

```
#View(NHANES)
attach(NHANES)
nhanes.data = data.frame(BPSysAve , Age, Height, Weight, Gender)
colSums(is.na(nhanes.data))
```

```
## BPSysAve      Age   Height   Weight   Gender
##     1449        0      353       78        0
```

```
n = nrow(nhanes.data)
```

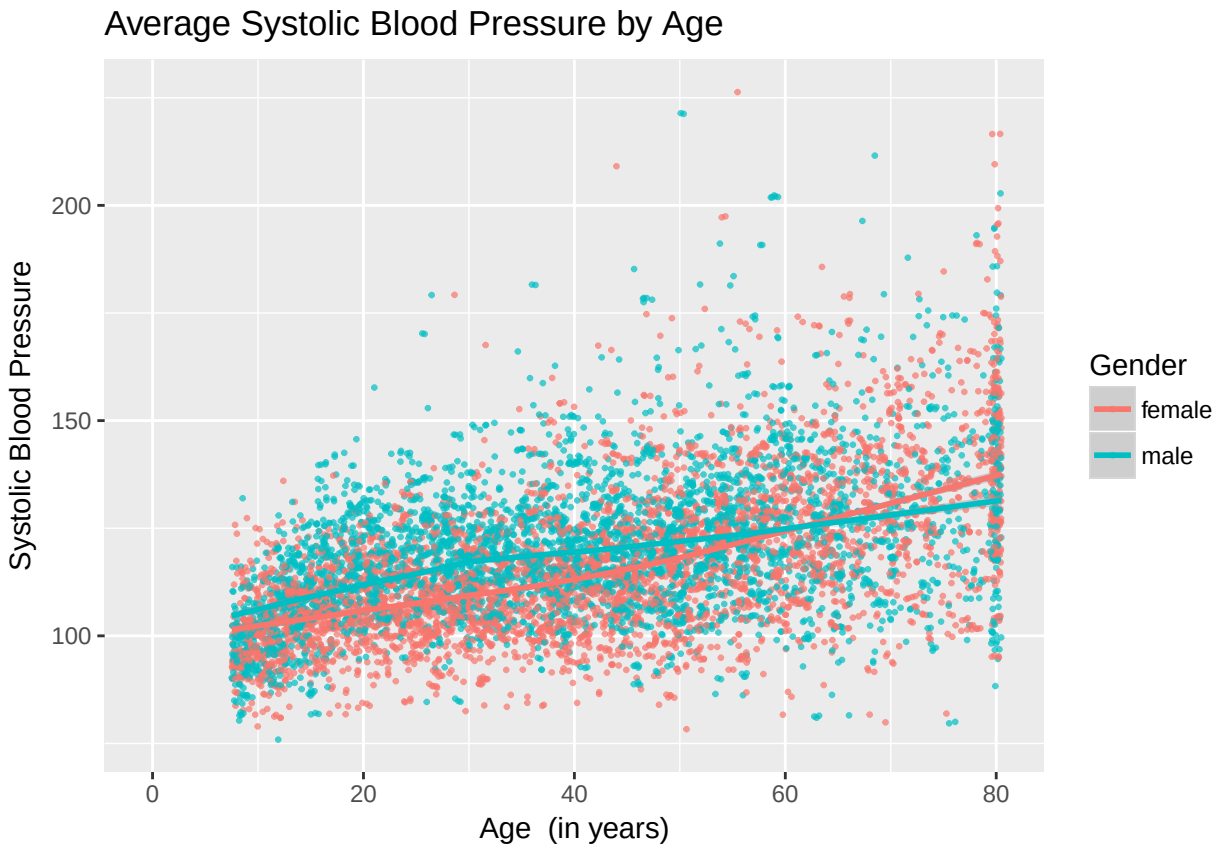
We have 1449 missing values of BPSysAve, 353 missing values of Height and 78 missing values of Weight. Since our dataset is large, we can ignore them.

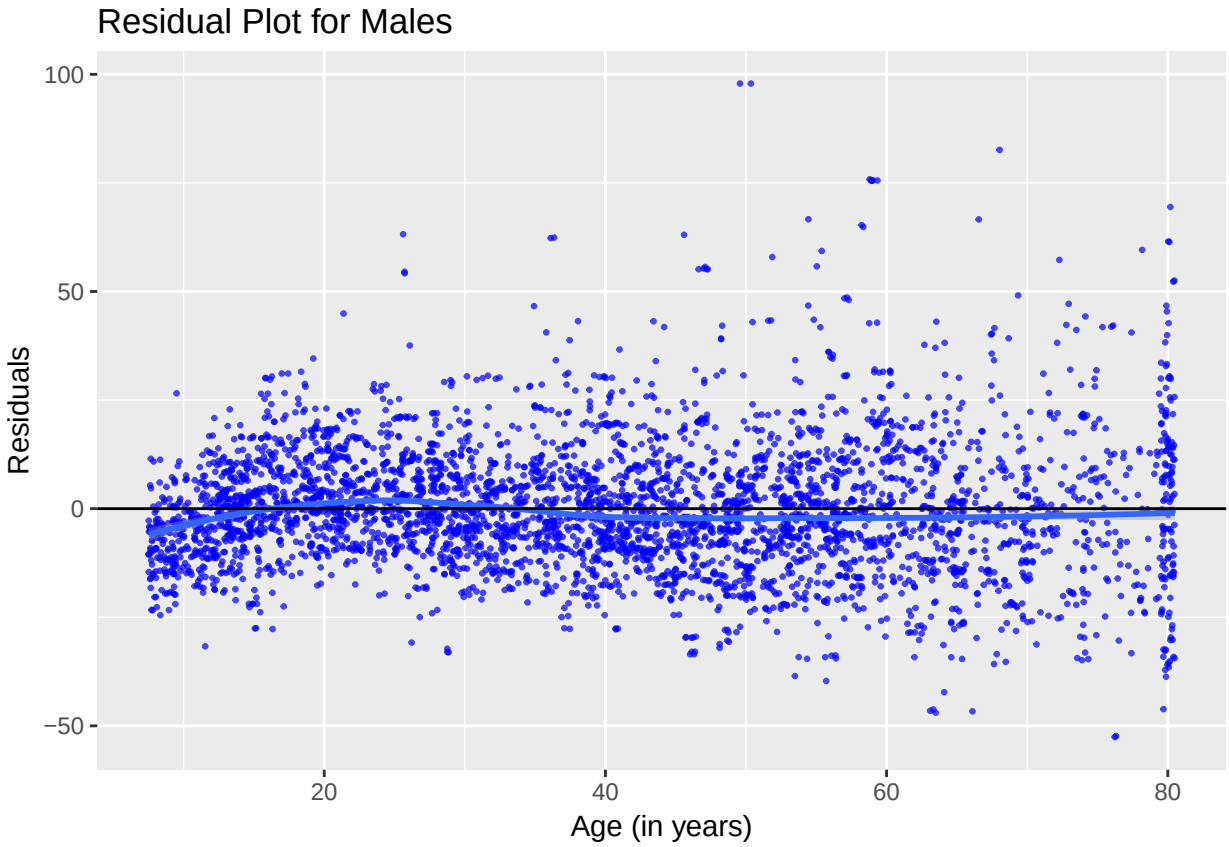
Let's try to find the relationship of average systolic blood pressure with age first

Section 01 - Relationship between Age vs Average Systolic Blood Pressure

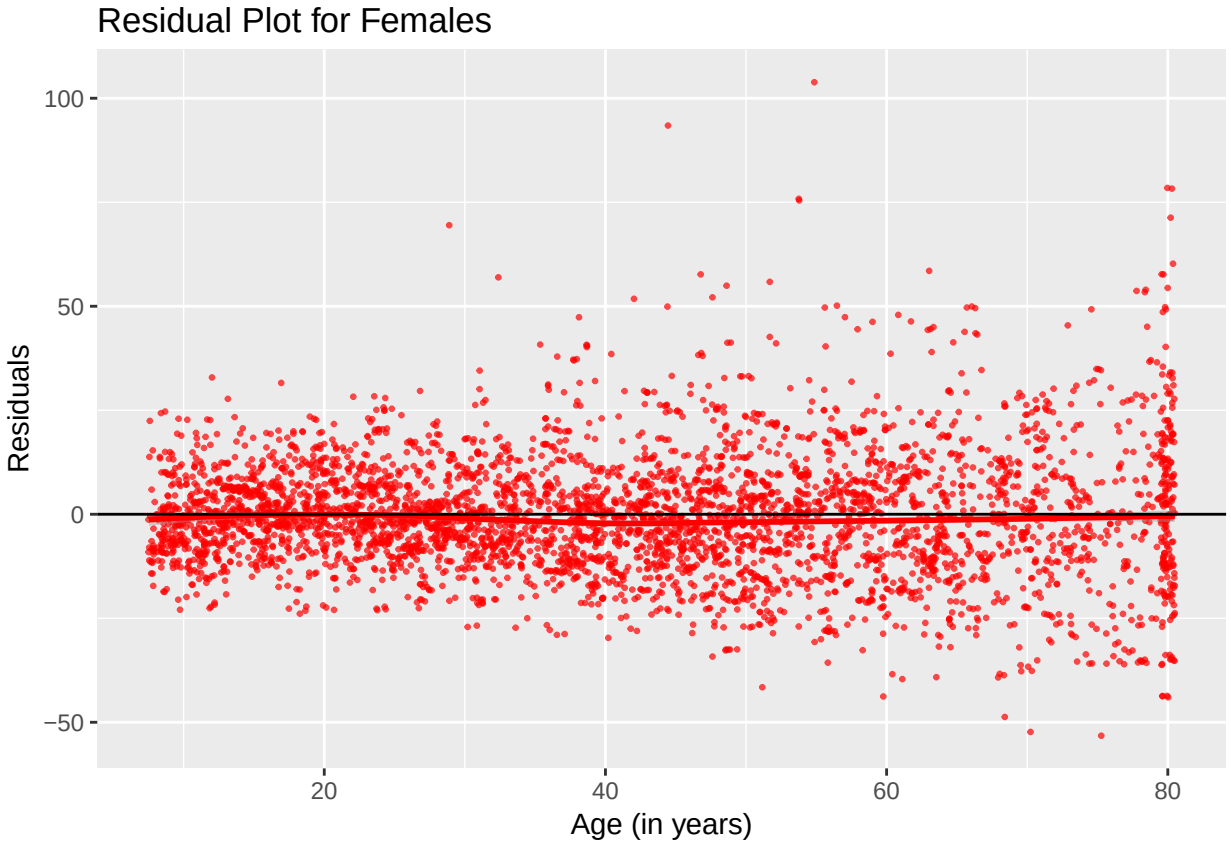
```
## Warning: Removed 1449 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1449 rows containing missing values (geom_point).
```


Average Systolic Blood Pressure by Age
Systolic Blood Pressure
200
150
100
0
20
40
60
80
Age (in years)
Gender
female
male


Residual Plot for Males
Residuals
100
50
0
−50
20
40
60
80
Age (in years)

```
## [1] 0.2014207
```



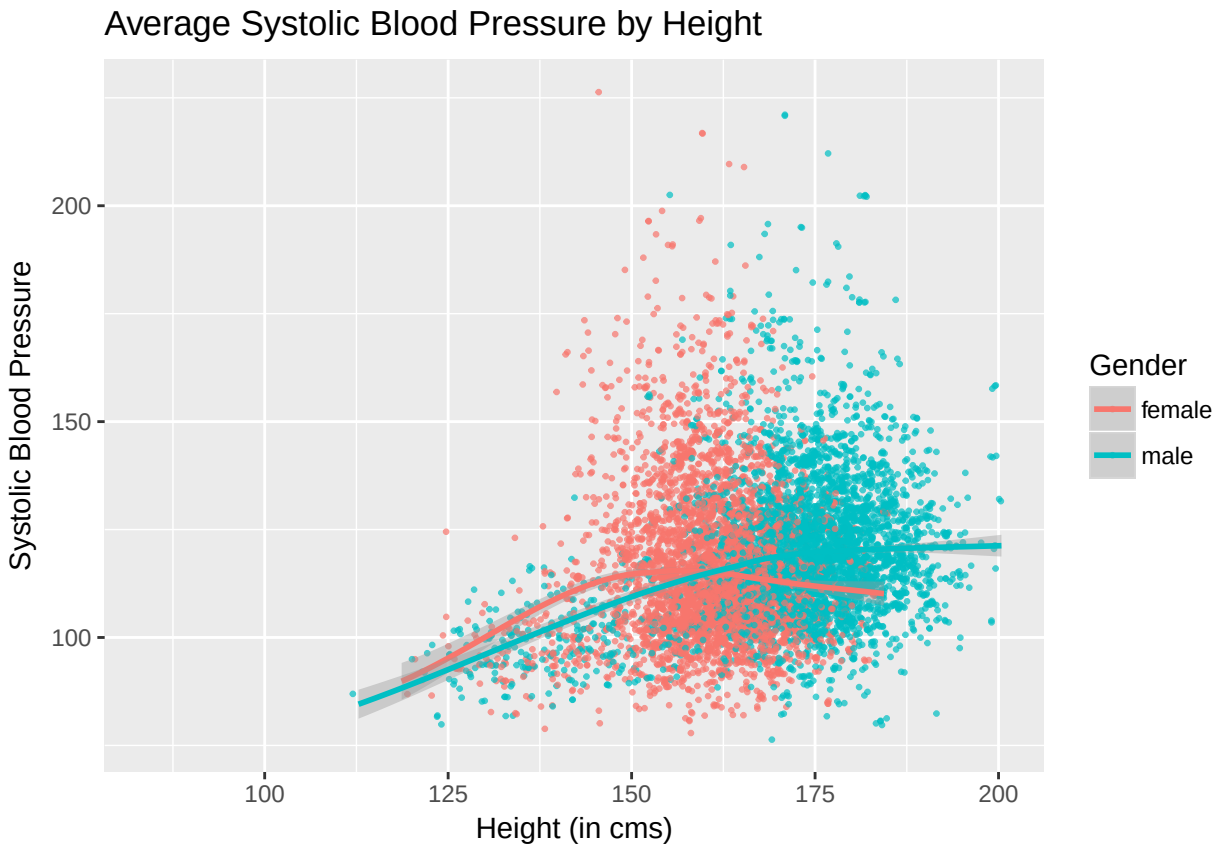

```
## [1] 0.3525604
```

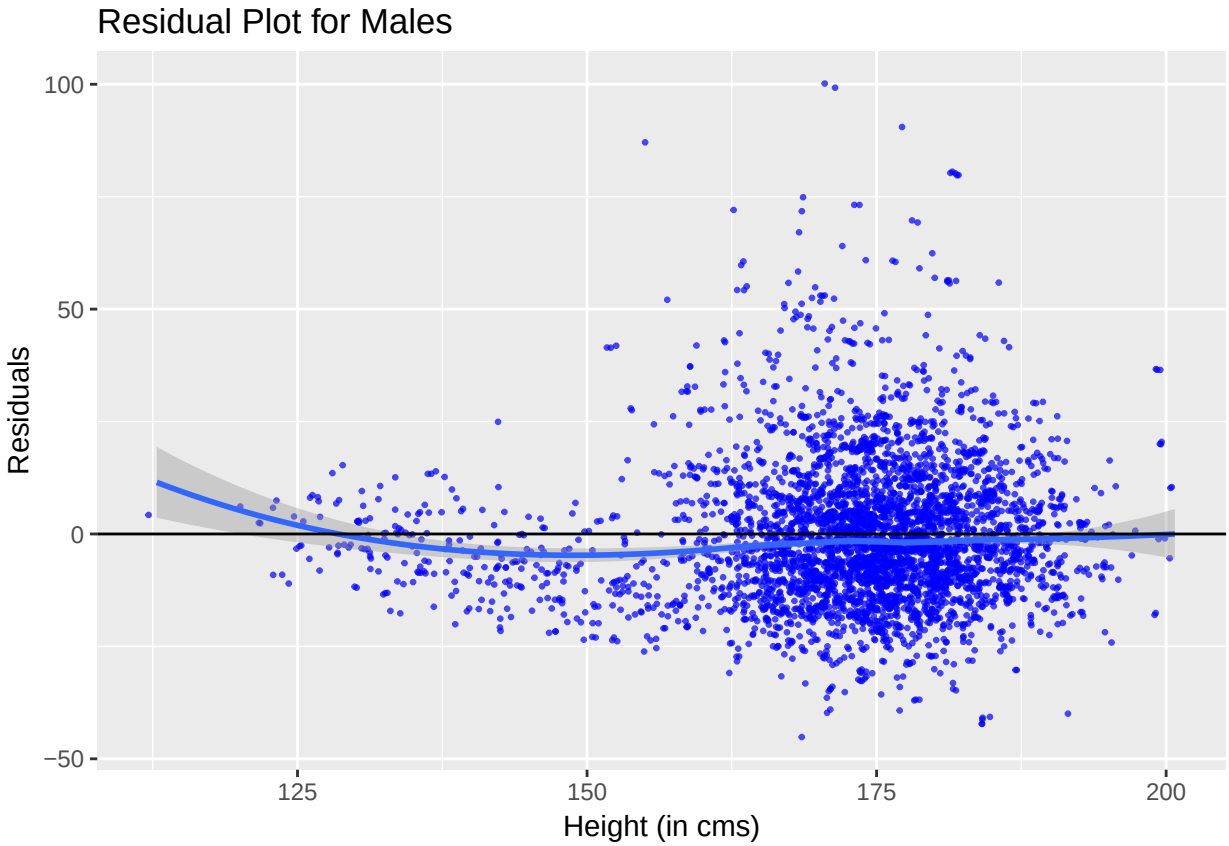
Both the male and female heights are right-skewed with 353 missing values. Since, the data is right-skewed we took median of the heights to handle missing values. The data seems to be rounded, so we unround it by adding random noise.

Section 02 - Relationship between Height vs Average Systolic Blood Pressure

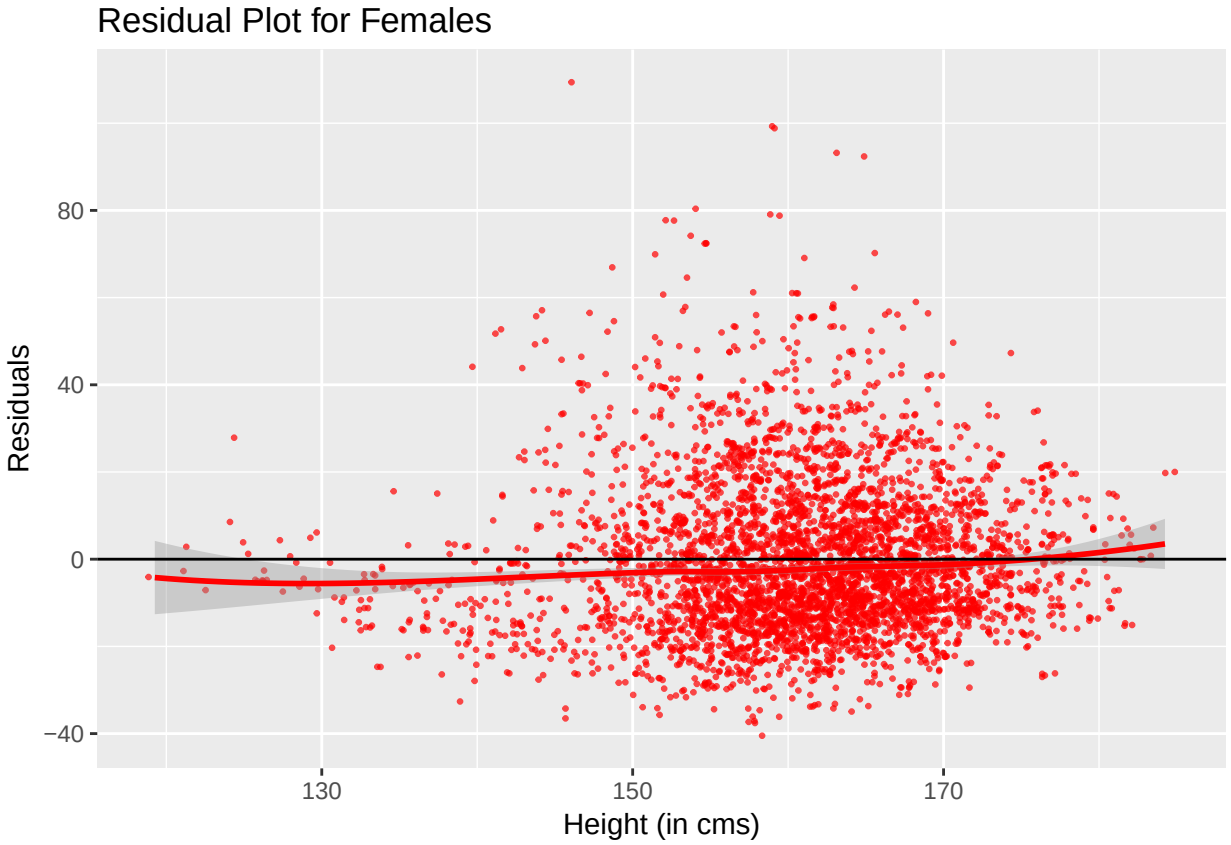
```
## Warning: Removed 1501 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1501 rows containing missing values (geom_point).
```


Average Systolic Blood Pressure by Height
Systolic Blood Pressure
200
150
100
100
125
150
175
200
Height (in cms)
Gender
female
male


Residual Plot for Males
Residuals
100
50
0
−50
125
150
175
200
Height (in cms)

```
## [1] 0.07056841
```



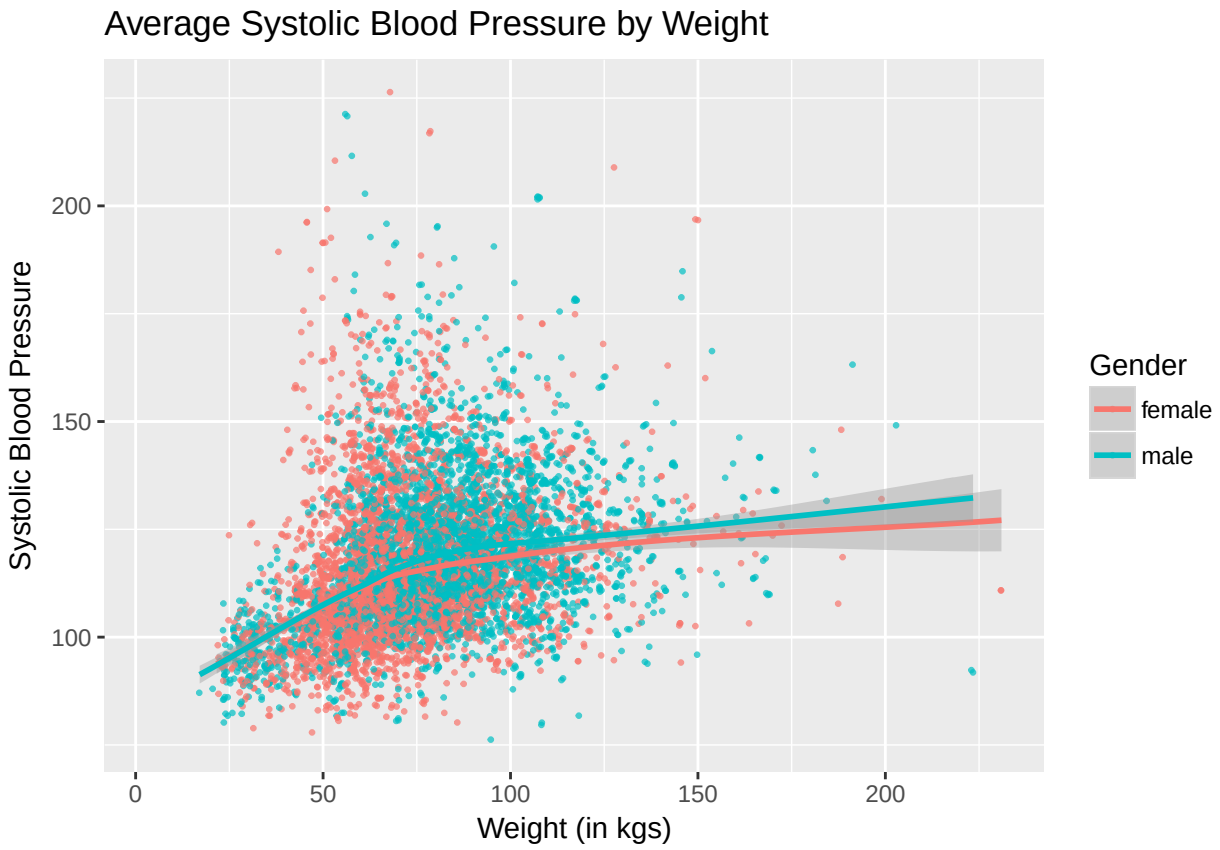

```
## [1] 0.02297047
```

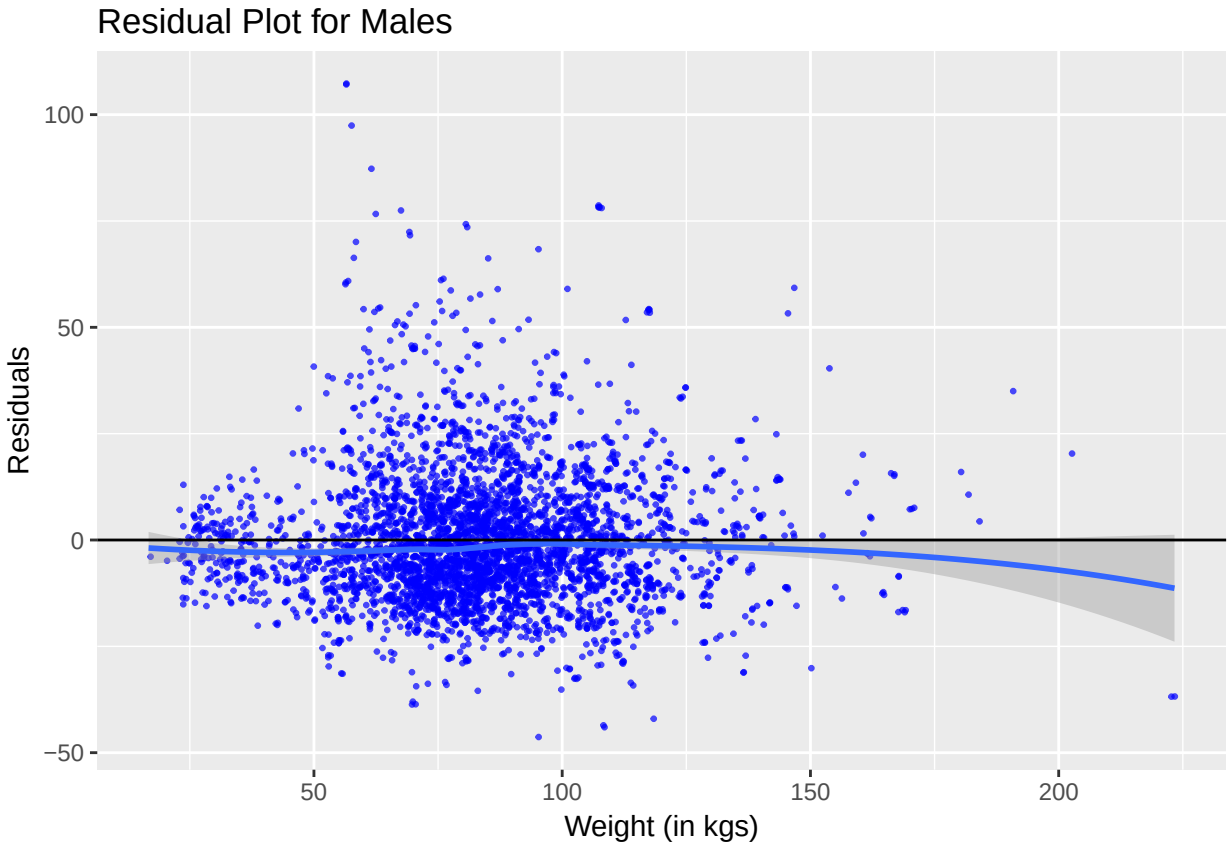
Both the male and female weights are left-skewed with 78 missing values. Since, the data is right-skewed we took median of the weights to handle missing values. 7% variation for males

Section 03 - Relationship between Weight vs Average Systolic Blood Pressure

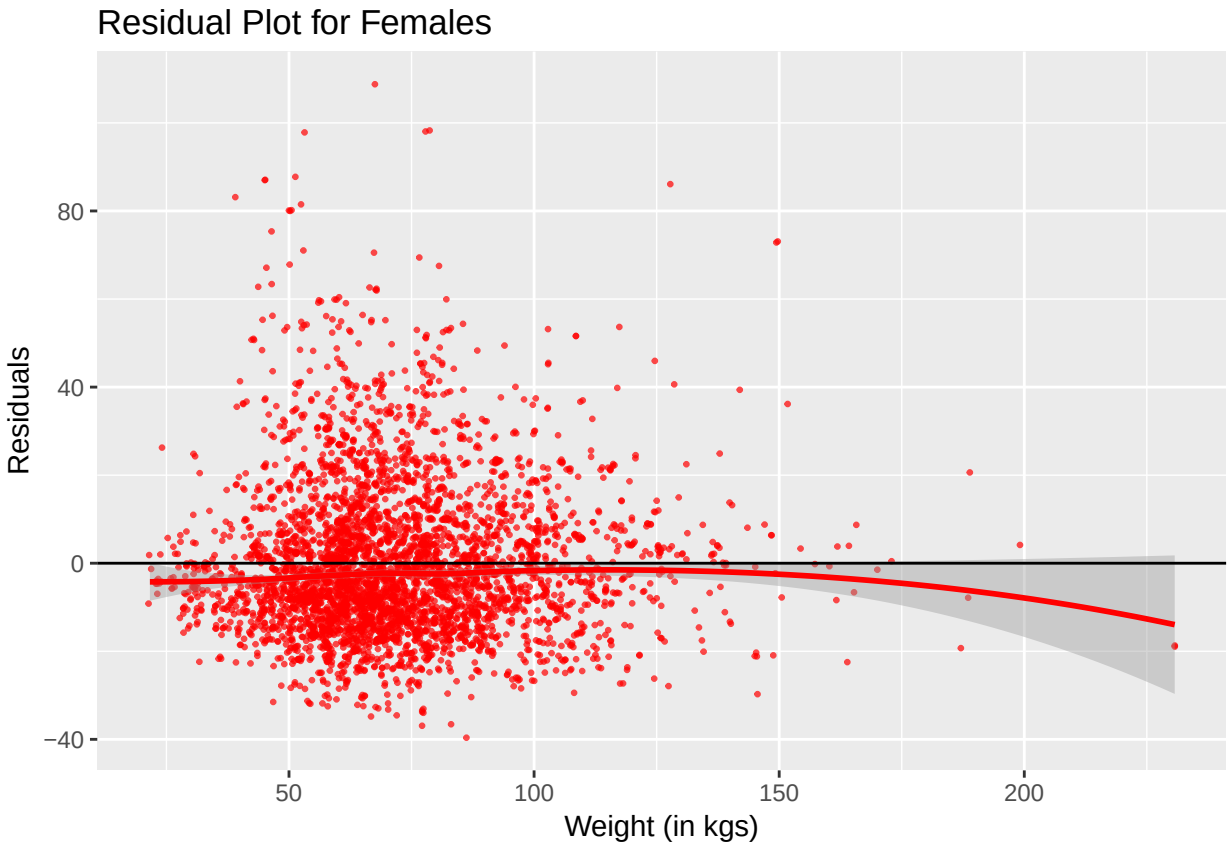
```
## Warning: Removed 1507 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1507 rows containing missing values (geom_point).
```


Average Systolic Blood Pressure by Weight
Systolic Blood Pressure
200
150
100
0
50
100
150
200
Weight (in kgs)
Gender
female
male


Residual Plot for Males
Residuals
100
50
0
−50
50
100
150
200
Weight (in kgs)

```
## [1] 0.1135723
```




```
## [1] 0.0627168
```

CONCLUSION