Problem Set 3

Sumeet Mishra

1/27/2019

Discussed with Kimball Wu,Pei-Yi-Cheng

```r
library(NHANES)
library(ggplot2)
library(broom)
library(arm)
nhanes<-(NHANES)

myvars1<-c("BPSysAve","Age","Gender")
myvars2<-c("BPSysAve","Height","Gender")
myvars3<-c("BPSysAve","Weight","Gender")

nhanes_age<-nhanes[myvars1]
nhanes_age$Age[ nhanes_age$Age > 80 ] = 80
nhanes_height<-nhanes[myvars2]
nhanes_weight<-nhanes[myvars3]
nhanes_age=nhanes_age[!(!complete.cases(nhanes_age)), ]
nhanes_height=nhanes_height[!(!complete.cases(nhanes_height)), ]
nhanes_weight=nhanes_weight[!(!complete.cases(nhanes_weight)), ]
```
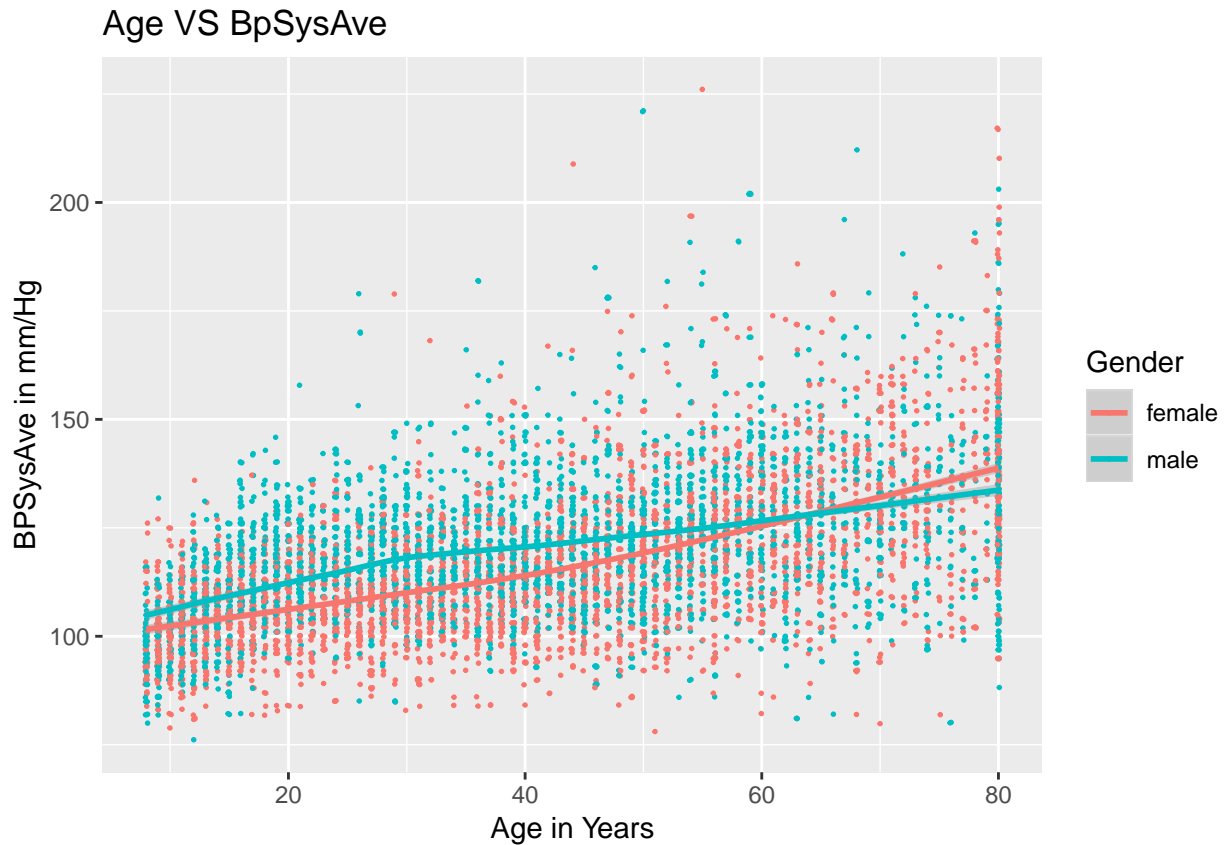
The NHANES library was loaded and the data is segregated into 3 models for Age,Height and Weight. For age model, if age found more than 80 then it was replaced with a value of 80. All the missing values were removed from the indivisual models.

Section-1(Age~BPSysAve)

```r
ggplot(nhanes_age, aes(x = Age, y =BPSysAve,color=Gender)) + geom_jitter(height = 0.2, width = 0.1, size
```

Age VS BpSysAve

The main plot for 'Age~BPSysAve' was plotted. The findings from the plot-

1.The systolic blood pressure has an increasing trend for both males and females.

2.Males has a slightly more systolic blood pressure than females of the same age. However, at around age 65, the pressure of both males and females are almost same and for afterwards the pressure for females increased more than males.

```
nhanes_age.lm = lm(BPSysAve ~ Age+Gender, data = nhanes_age)
nhanes_age.lm.df = augment(nhanes_age.lm)
ggplot(nhanes_age.lm.df, aes(x =Age , y = .resid)) + geom_jitter(height = 0.3, width = 0.1, size = 0.3)
```

## Residual graph of Age with respect to BPSysAve
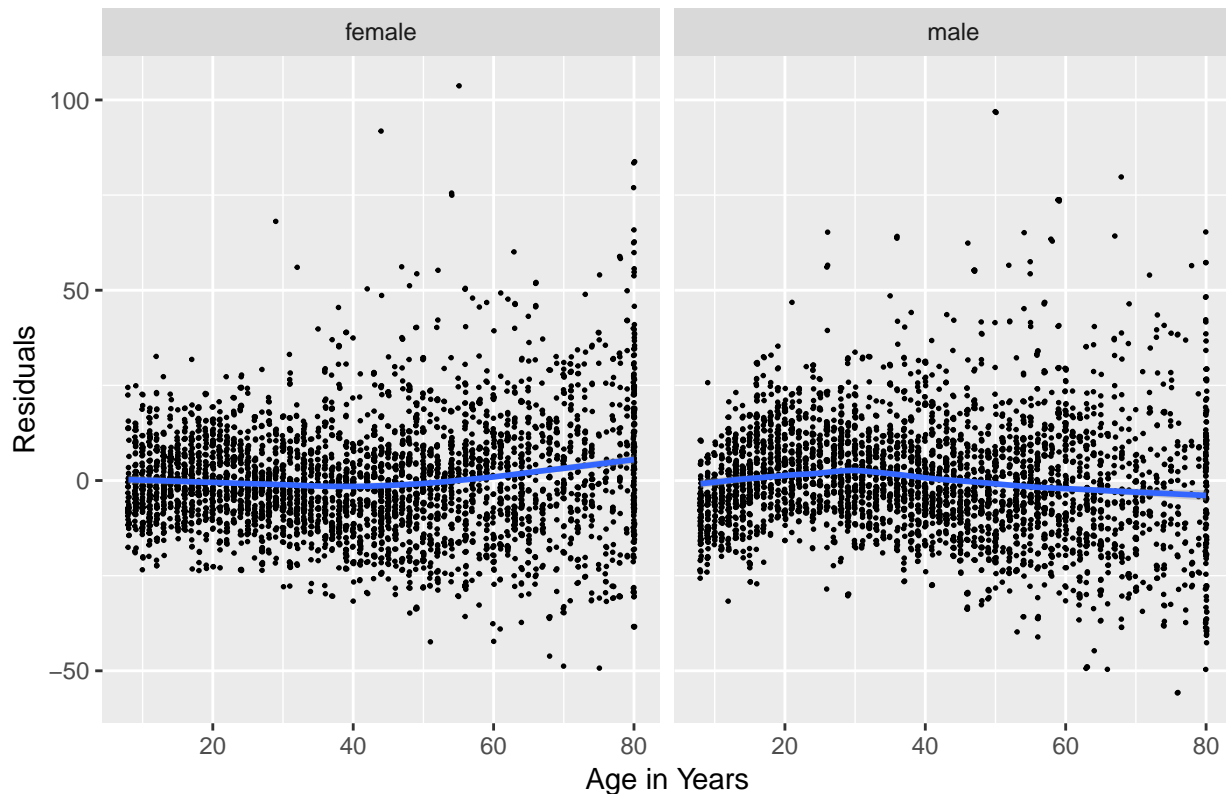

```
#display(nhanes_age.lm)
```

From the residual plot for age we observed the trend on a linear model. The model describes and captures the trend well. So, we can conclude that linear model is a good model for this data. (Although other models may be good or better than linear model)

Section-2(Height~BPSysAve)

```
ggplot(nhanes_height, aes(x = Height, y =BPSysAve,color=Gender)) + geom_jitter(height = 0.3, width = 0.
```
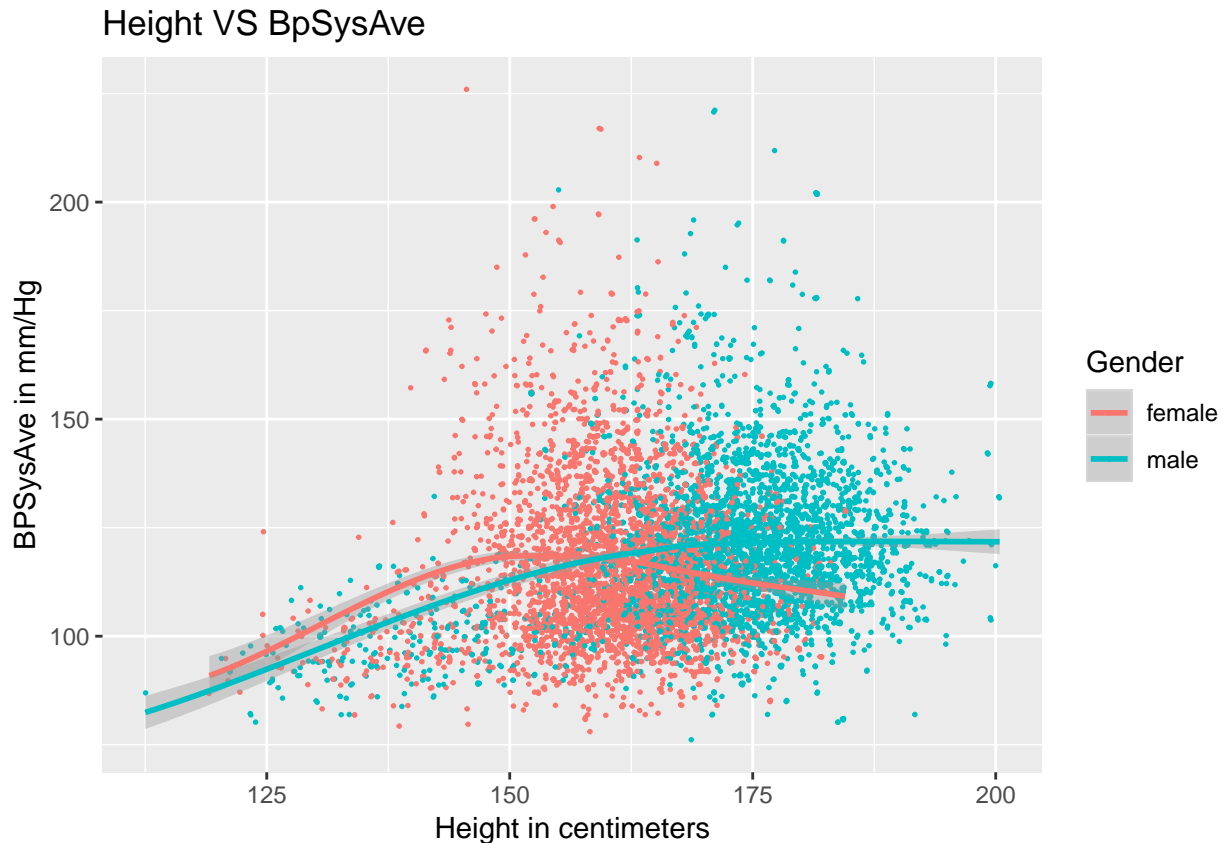
Height VS BpSysAve

The plot between 'Height~BPSysAve' was plotted. That shows at around the height of 163 centimeters there was an intersection between males and females. Previous to that point there is an increasing trend between males and females,females having a larger blood pressure than males .But,afterwards it was observed that the blood pressure decreased a lot for females where as for males it keeps increasing. The initial lower value of both blood pressure and height may be assumed as children.

```
nhanes_height.lm = lm(BPSysAve ~ Height + Gender, data = nhanes_height)

nhanes_height.lm.df = augment(nhanes_height.lm)

ggplot(nhanes_height.lm.df, aes(x =Height , y = .resid)) + geom_jitter(height = 0.3, width = 0.1, size =
  facet_wrap(~ Gender) + geom_smooth(method='loess',method.args=list(degree=1))+xlab('Height in centimet
```

Residual graph of Height with respect to BPSysAve
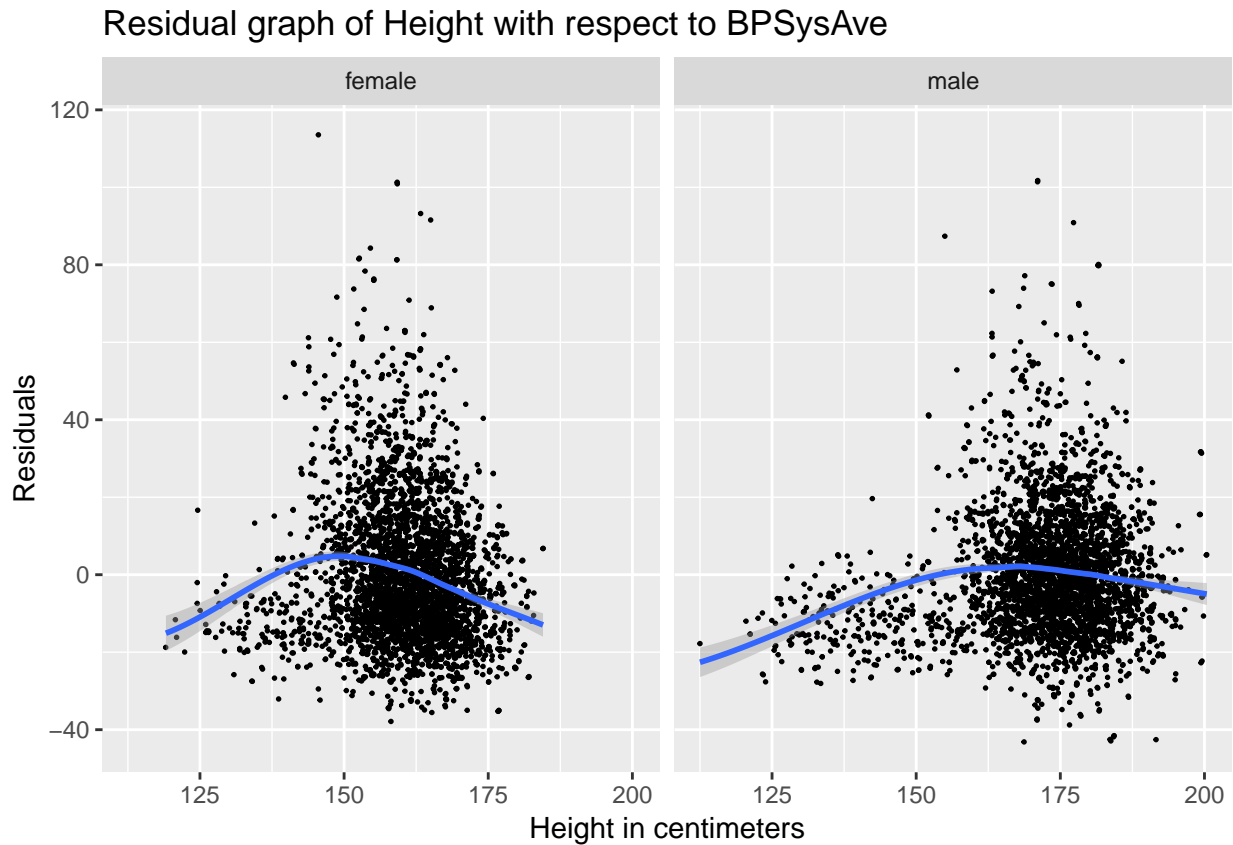
```
#display(nhanes_height.lm)
```

The residual plot for Height is drawn with linear model. It can be seen that linear model does not describe well the plot. But other models like sqaure model were tried which does not improve much the graph. Which means height alone is not a very good predictor of blood pressure.

Section-3(Weight~BPSysAve)

```
ggplot(nhanes_weight, aes(x = Weight, y =BPSysAve,color=Gender)) + geom_point()  + geom_jitter(height =
```
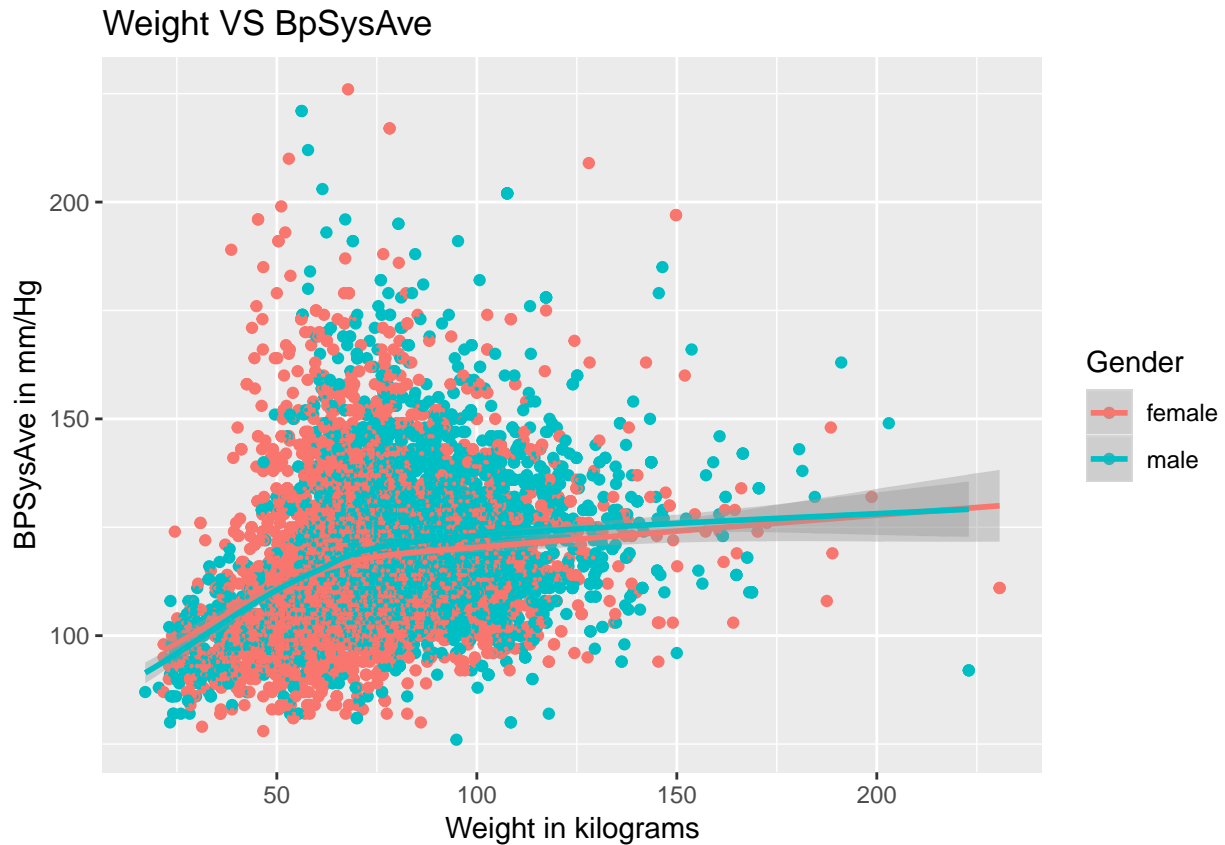
Weight VS BpSysAve

The plot between 'Weight~ BPSysAve' was plotted. From the plot, it was observed that upto weight 75kg, the blood pressure increased a lot but afterwards it increased at steady pace. The graph does not show any significant difference between males and females.

```
nhanes_weight.lm = lm(BPSysAve ~ log(Weight) + Gender+Weight, data = nhanes_weight)
nhanes_weight.lm.df = augment(nhanes_weight.lm)
ggplot(nhanes_weight.lm.df, aes(x =Weight , y = .resid)) + geom_jitter(height = 0.3, width = 0.1, size =
  facet_wrap(~ Gender) + geom_smooth(method='loess',method.args=list(degree=1))+xlab('log(Weight in kilo
```

Residual graph of Weight with respect to BPSysAve
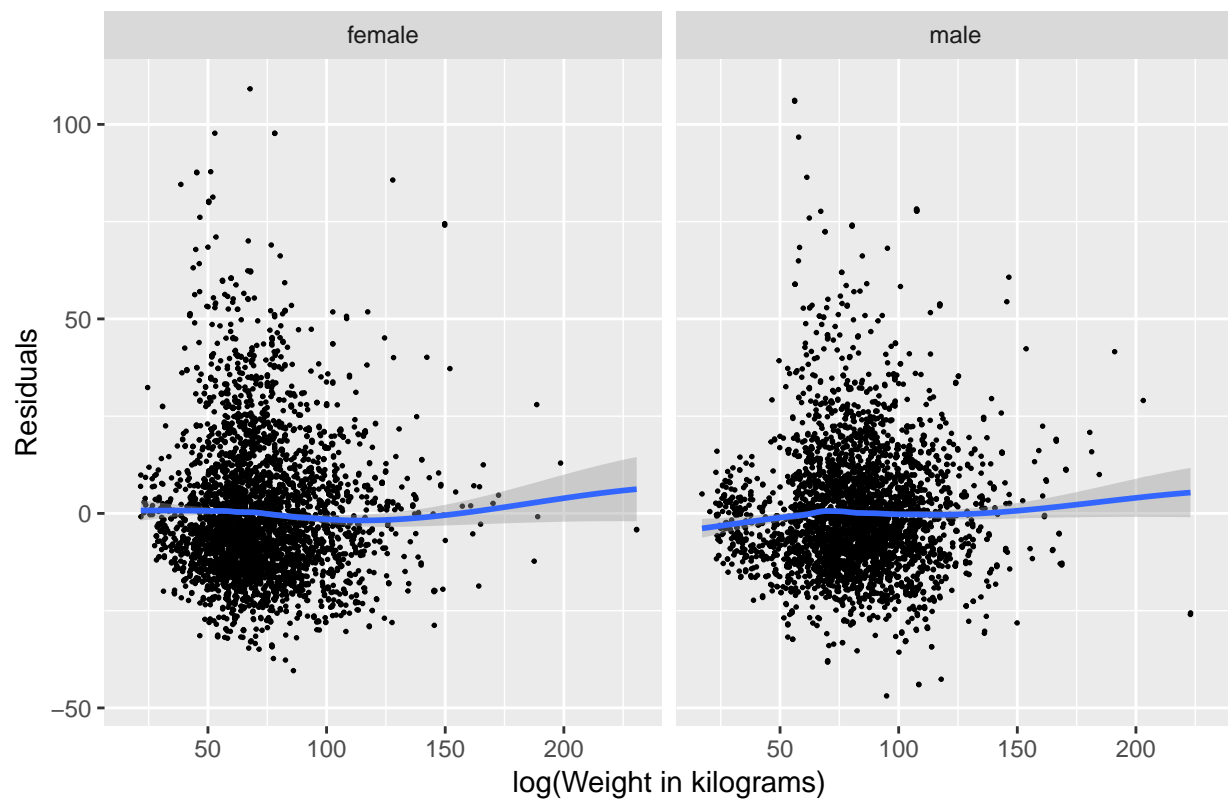

```
#display(nhanes_weight.lm)
```

The residual plot for Weight is drawn with linear model. Since we observed an increasing trend in the 'weight~BpSysAve', we have taken log(Weight) which improved r squared value,although not significantly. This is because weight alone is not a good predictor of blood pressure as it is weakly correlated with blood pressure.