CHAPTER 1

Andaleeb Hassan

2025-08-17

1 Inheritance of Height

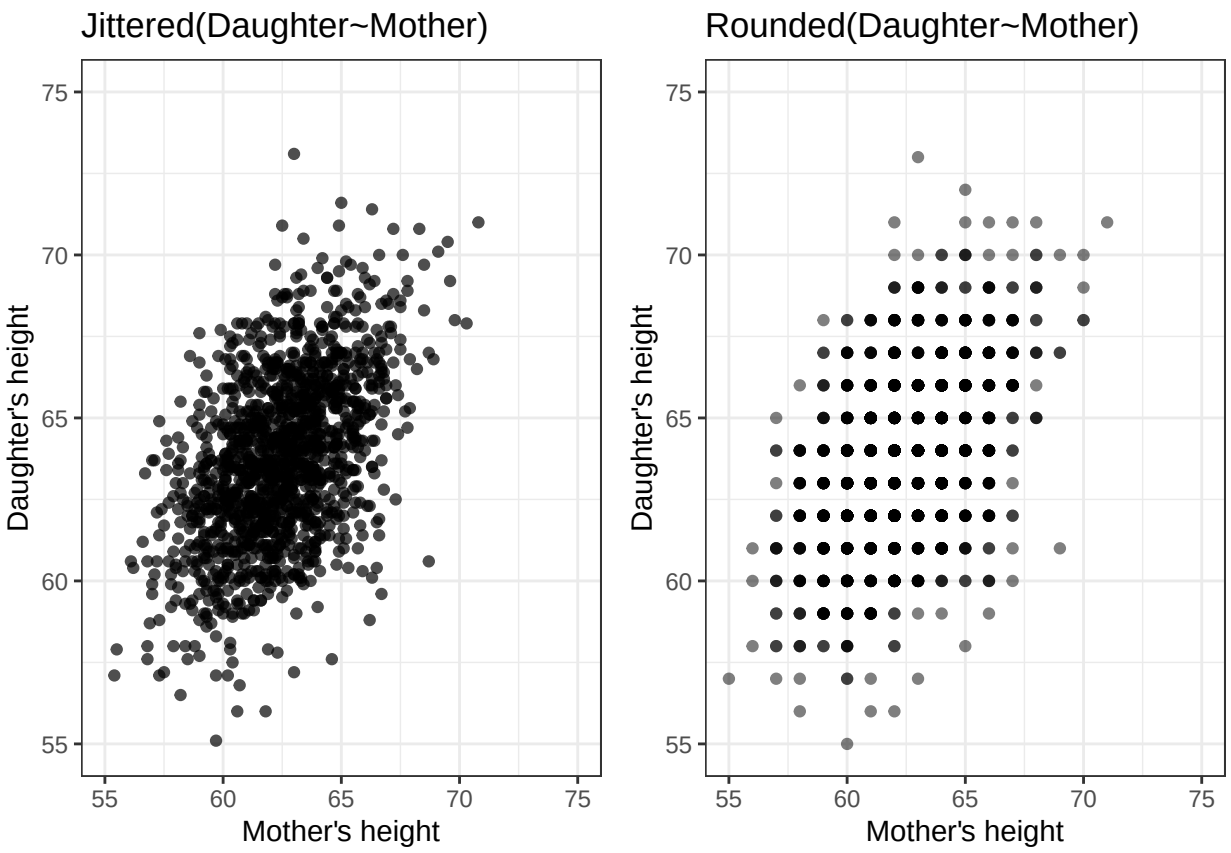
```
h_1 <- tibble(mh2 = round(Heights$mheight),dh2 = round(Heights$dheight),Heights)

h_1 %>% head()
```

```
## # A tibble: 6 x 4
##     mh2   dh2 mheight dheight
##   <dbl> <dbl>   <dbl>   <dbl>
## 1    60    55    59.7    55.1
## 2    58    56    58.2    56.5
## 3    61    56    60.6    56
## 4    61    57    60.7    56.8
## 5    62    56    61.8    56
## 6    56    58    55.5    57.9
```

1.1 Jittered and Rounded scatterplot

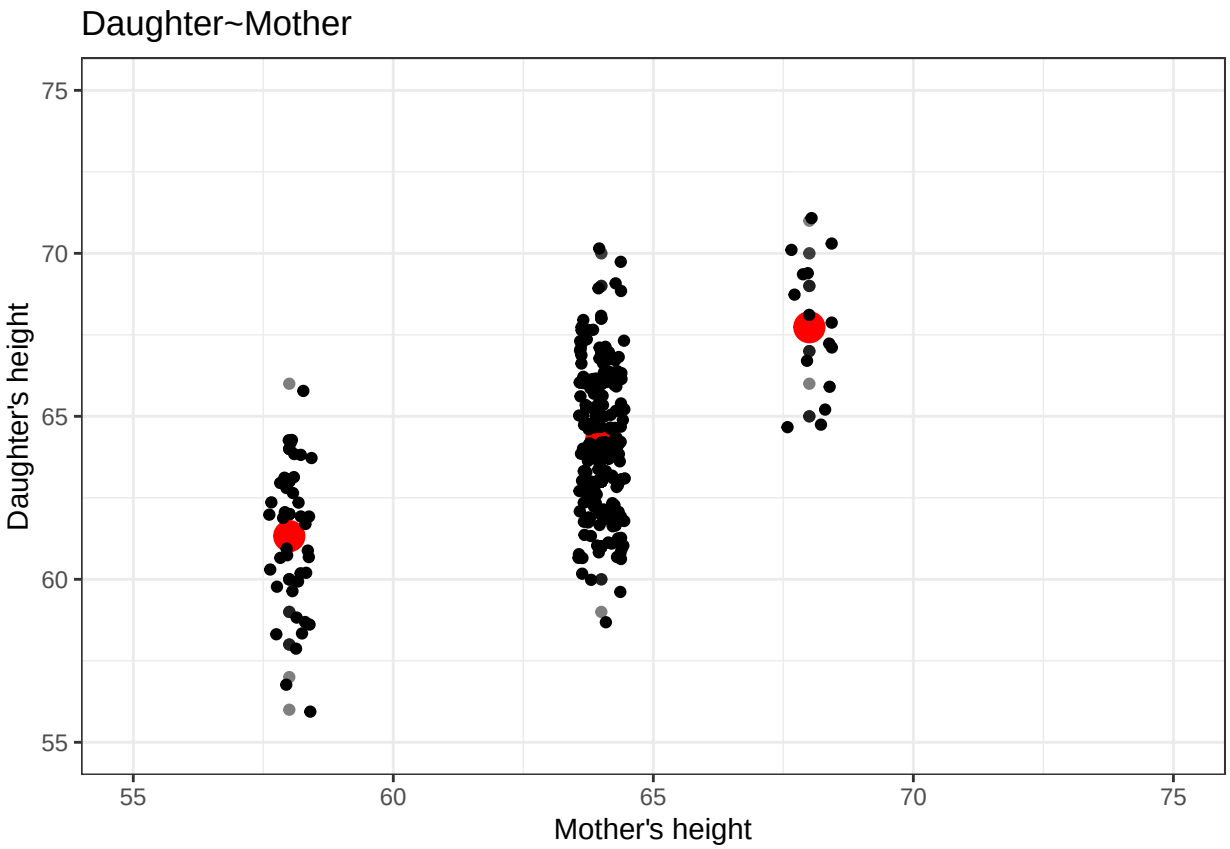
```
#code for drawing the graphs

M1 <- ggplot(data = Heights,mapping = aes(mheight,dheight))+
  geom_point(alpha = .7, col = "black")+
  ggtitle("Jittered(Daughter~Mother)")+
  xlab("Mother's height")+
  ylab("Daughter's height")+
  xlim(55,75)+
  ylim(55,75)+
  theme_bw()

M2 <- ggplot( data = h_1, mapping = aes(x = mh2, y =dh2))+
     geom_point(alpha = .5)+
  ggtitle("Rounded(Daughter~Mother)")+
  xlab("Mother's height")+
  ylab("Daughter's height")+
  xlim(55,75)+
  ylim(55,75)+
  theme_bw()
```

```
grid.arrange(M1,M2,ncol =2)
```



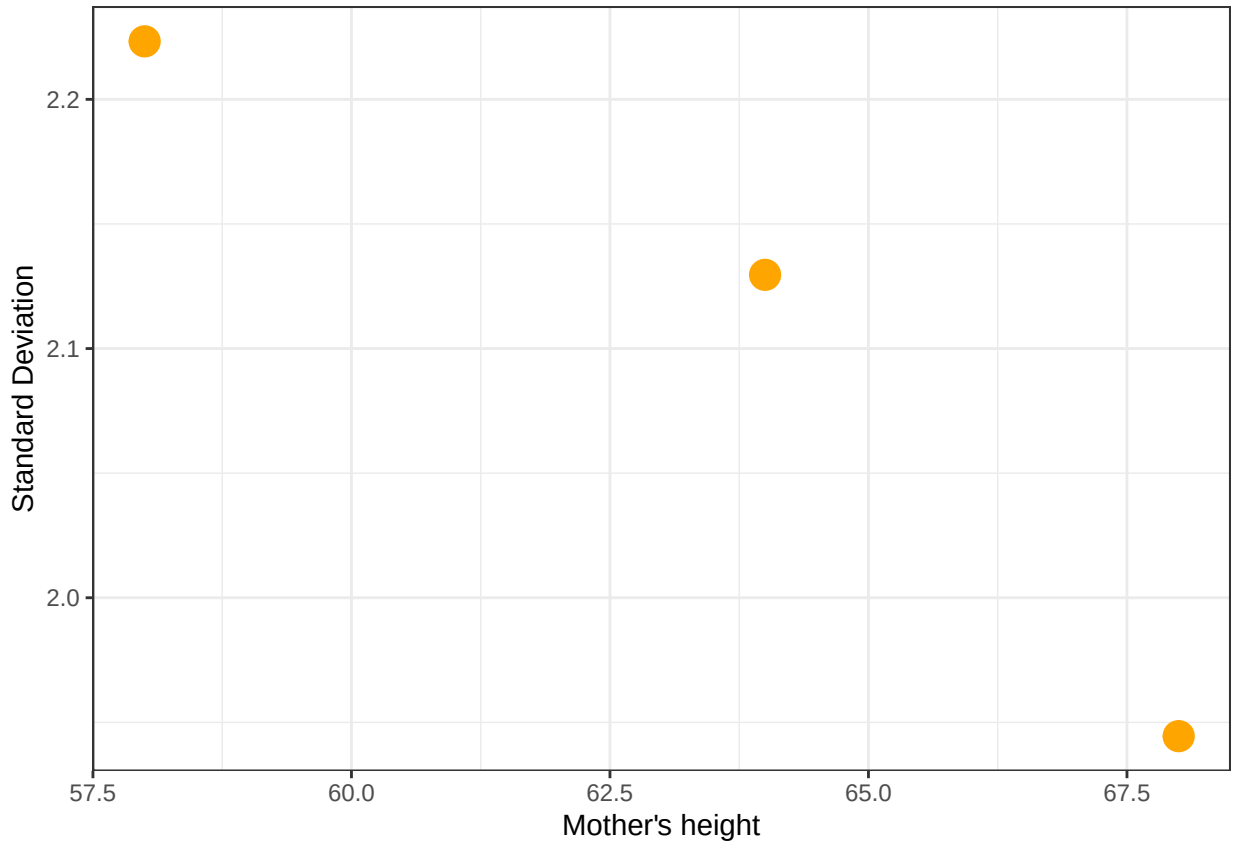
1.2 Stripwise

```
h_2 <- tibble(mhs = h_1$mh2[h_1$mh2 %in%c(58,64,68)],
              dhs = h_1$dh2[h_1$mh2 %in%c(58,64,68)])

   ggplot(data =h_2, mapping = aes(mhs,dhs))+
  geom_point(alpha = .5)+
  stat_summary(geom = "point",fun.y =  "mean",col ="red",size =5)+
  ggtitle("Daughter~Mother")+
  xlab("Mother's height")+
  ylab("Daughter's height")+
  xlim(55,75)+
  ylim(55,75)+
  geom_jitter(width = .45)+
  theme_bw()
```



1.2.1 variation among the strips

```
ggplot(data =h_2, mapping = aes(mhs,dhs))+
  stat_summary(geom = "point",fun.y = "sd",col ="orange",size =5)+
  xlab("Mother's height")+
  ylab("Standard Deviation")+
  theme_bw()
```




```
h_2%>%
  group_by(mhs)%>%
  summarise(average = mean(dhs),variability = sd(dhs))
```

```
## # A tibble: 3 x 3
##     mhs average variability
##   <dbl>   <dbl>       <dbl>
## 1    58    61.3        2.22
## 2    64    64.3        2.13
## 3    68    67.7        1.94
```

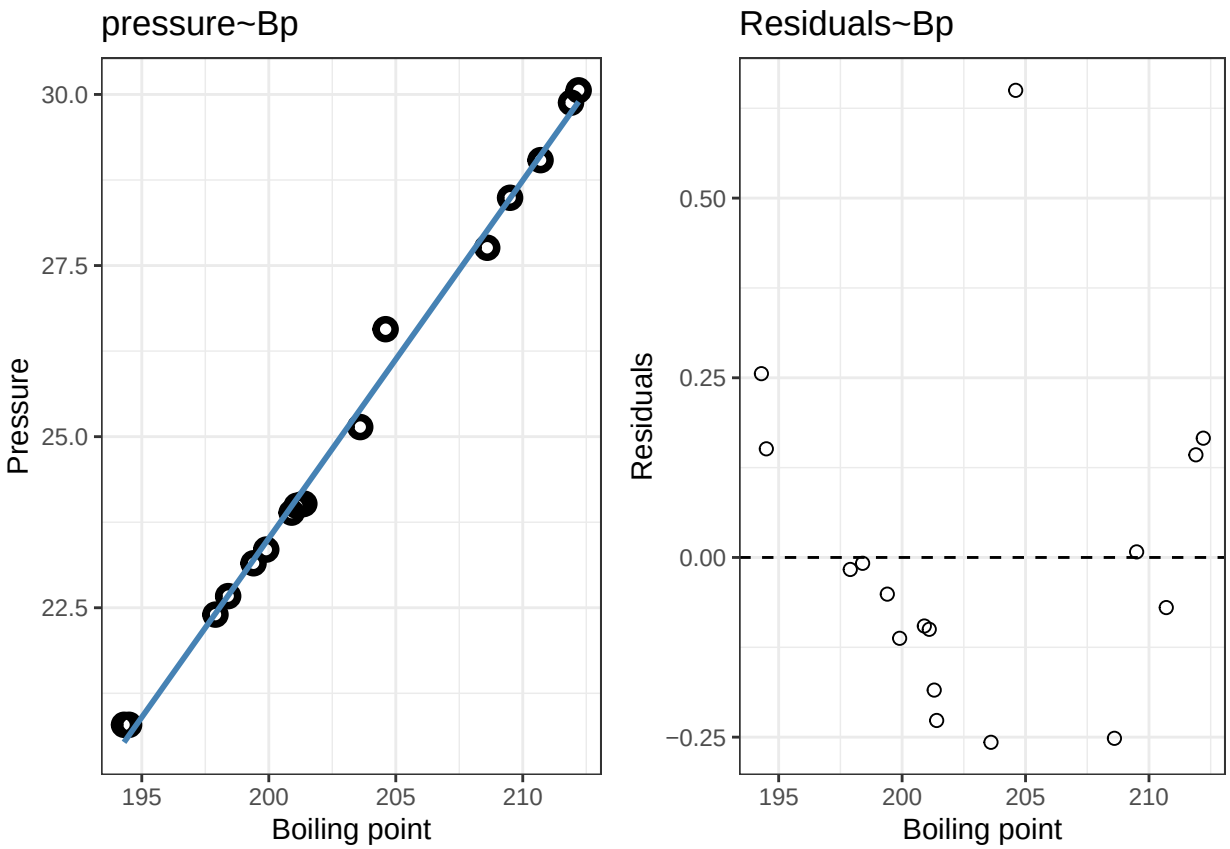
Here we can see that the mean is Rising from left to right.

2 Forbes's Data

2.1 pressure~bp (scatterplot and residual plot)

```
# Scatterplot and smooth curve

M3 <- ggplot(data = Forbes, mapping = aes(bp,pres))+
  geom_point(shape = 21,size = 2,stroke = 2)+
  ggtitle("pressure~Bp")+
  xlab("Boiling point")+
  ylab("Pressure")+
  geom_smooth(method ="lm",se =FALSE,col ="steelblue")+
  theme_bw()

# Residuals vs Boiling point

lm <- aov(Forbes$pres~Forbes$bp)

fr <- tibble(resid = lm$residuals,Forbes)

M4 <- ggplot(data = fr, mapping = aes(bp,resid))+
  geom_point(size =2,shape =21)+
  ggtitle("Residuals~Bp")+
  xlab("Boiling point")+
  ylab("Residuals")+
  geom_hline(yintercept=0,linetype=2)+
  theme_bw()
```

```
grid.arrange(M3,M4,ncol =2)
```

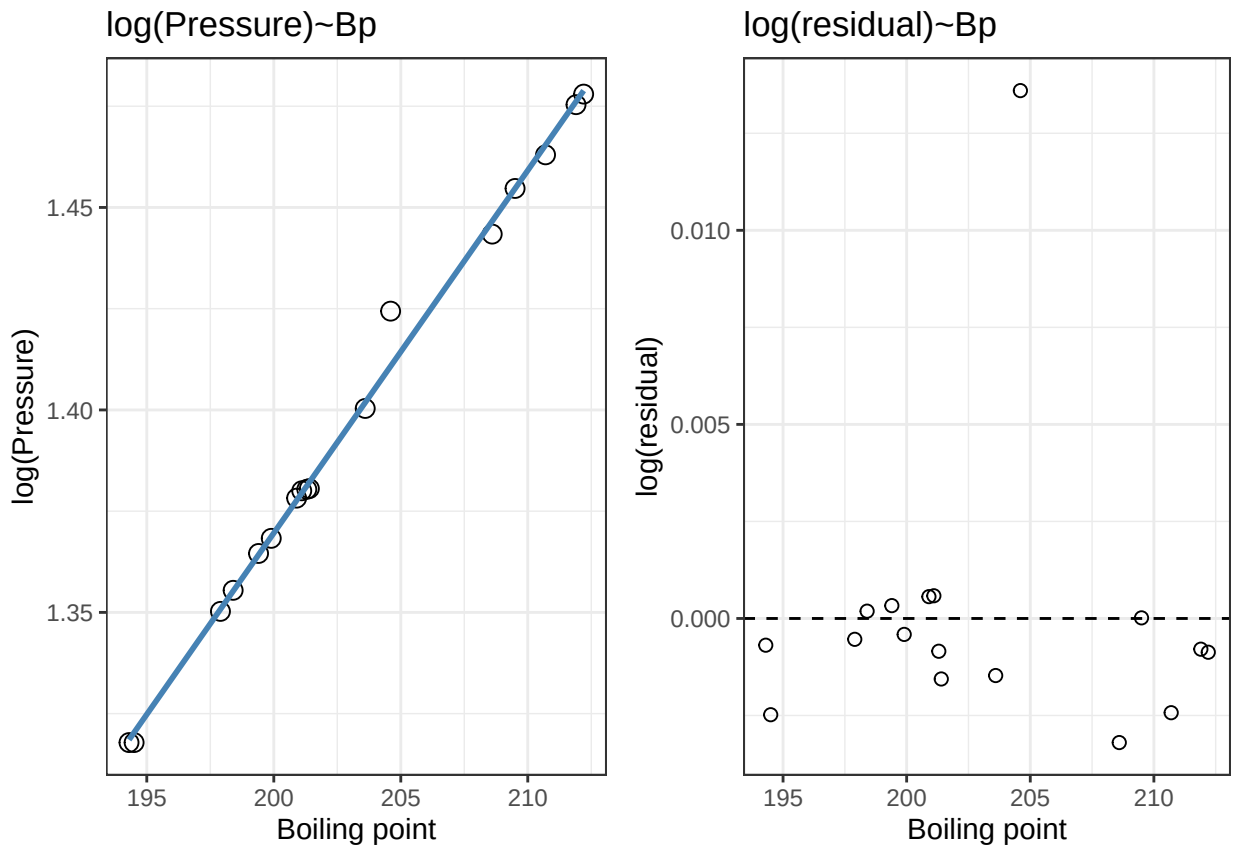


Forbes suggested that log(pressure) is linearly related to Boiling point.

2.2 log(pres)~bp(Scatterplot and residual plot)

```
fr2 <- tibble(Forbes,lnpres = log10(Forbes$pres))

ml2 <- aov(fr2$lnpres~fr2$bp)

fr3 <-tibble(lnpres = log10(Forbes$pres),lnres = ml2$residuals,Forbes)

M5 <- ggplot(data = fr2, mapping = aes(bp,lnpres))+
     geom_point(size =3,shape =21)+
     ggtitle("log(Pressure)~Bp")+
     xlab("Boiling point")+
     ylab("log(Pressure)")+
     geom_smooth(method ="lm",se =FALSE,col ="steelblue")+
     theme_bw()

M6 <- ggplot(data = fr3, mapping = aes(bp,lnres))+
    geom_point(size =2,shape =21)+
    ggtitle("log(residual)~Bp")+
    xlab("Boiling point")+
    ylab("log(residual)")+
```

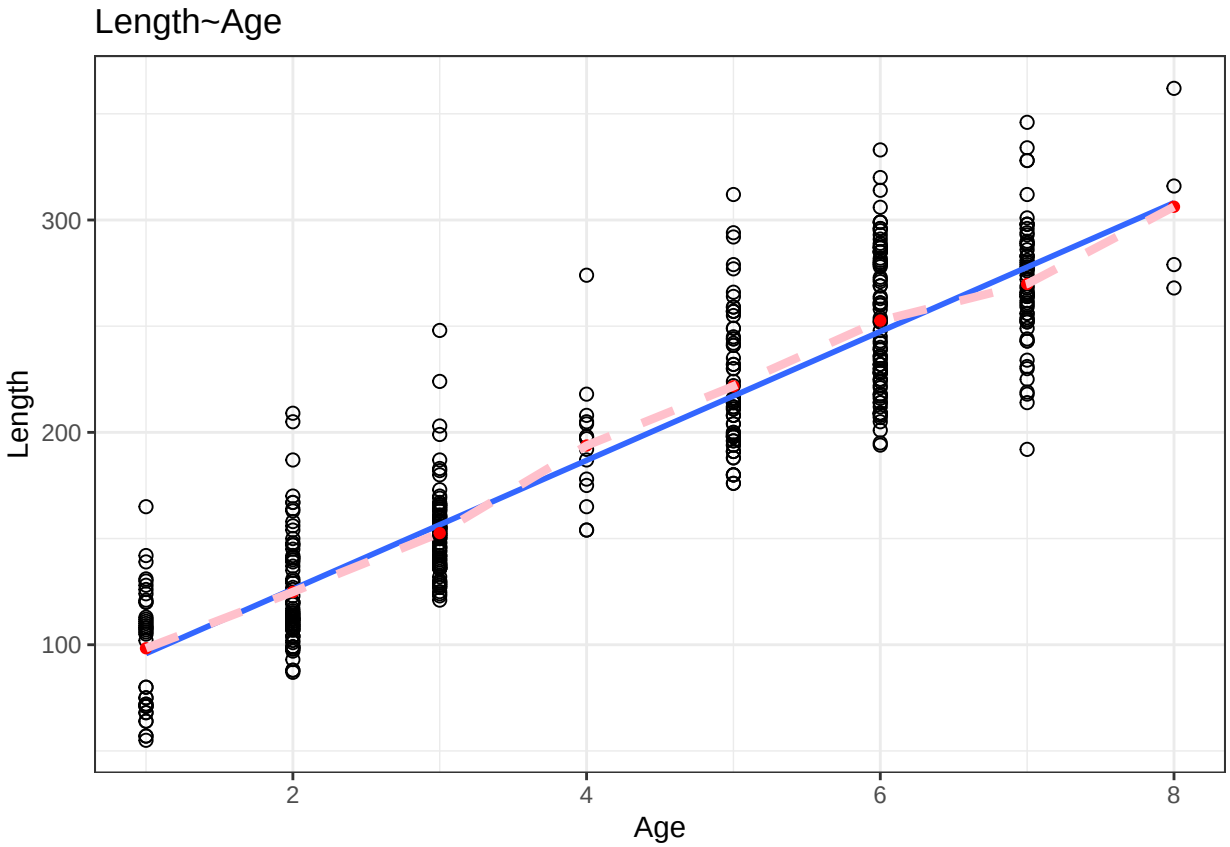
```
  geom_hline(yintercept=0,linetype=2)+
  theme_bw()

grid.arrange(M5,M6,ncol =2)
```



3 Length at age for Smallmouth Bass

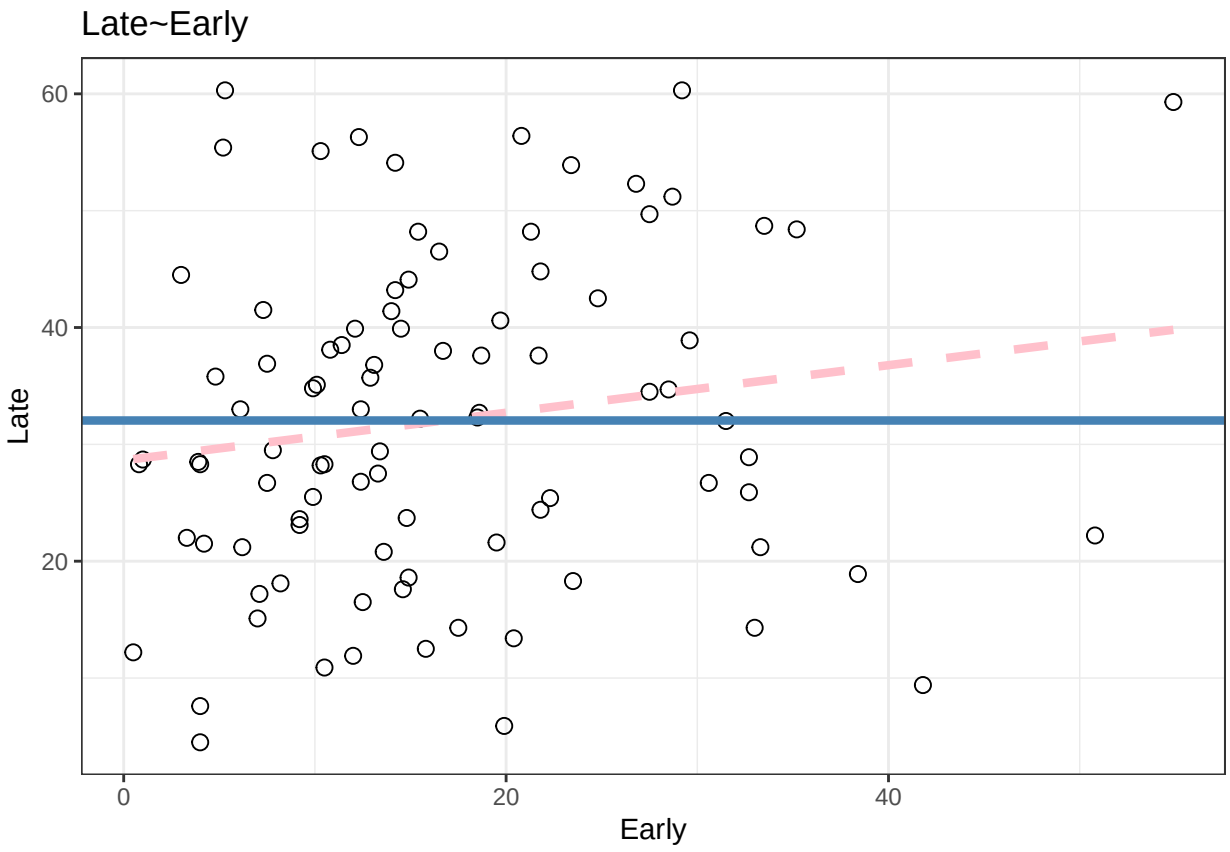
```
avg <- wblake%>%
  group_by(Age)%>%
  summarise(mean = mean(Length))
mpoint <- tibble(avg$age,avg$mean)
```

```
ggplot( data = wblake,mapping = aes(Age,Length))+
  geom_point( shape = 21, size =2)+
  ggtitle("Length~Age")+
  geom_smooth(se = FALSE,method = "lm")+
  stat_summary(geom = "point",fun.y = "mean",col ="red")+
  stat_summary(geom =  "line",fun.y = "mean",col = "pink",size = 1.5,
               linetype = "dashed")+
  theme_bw()
```



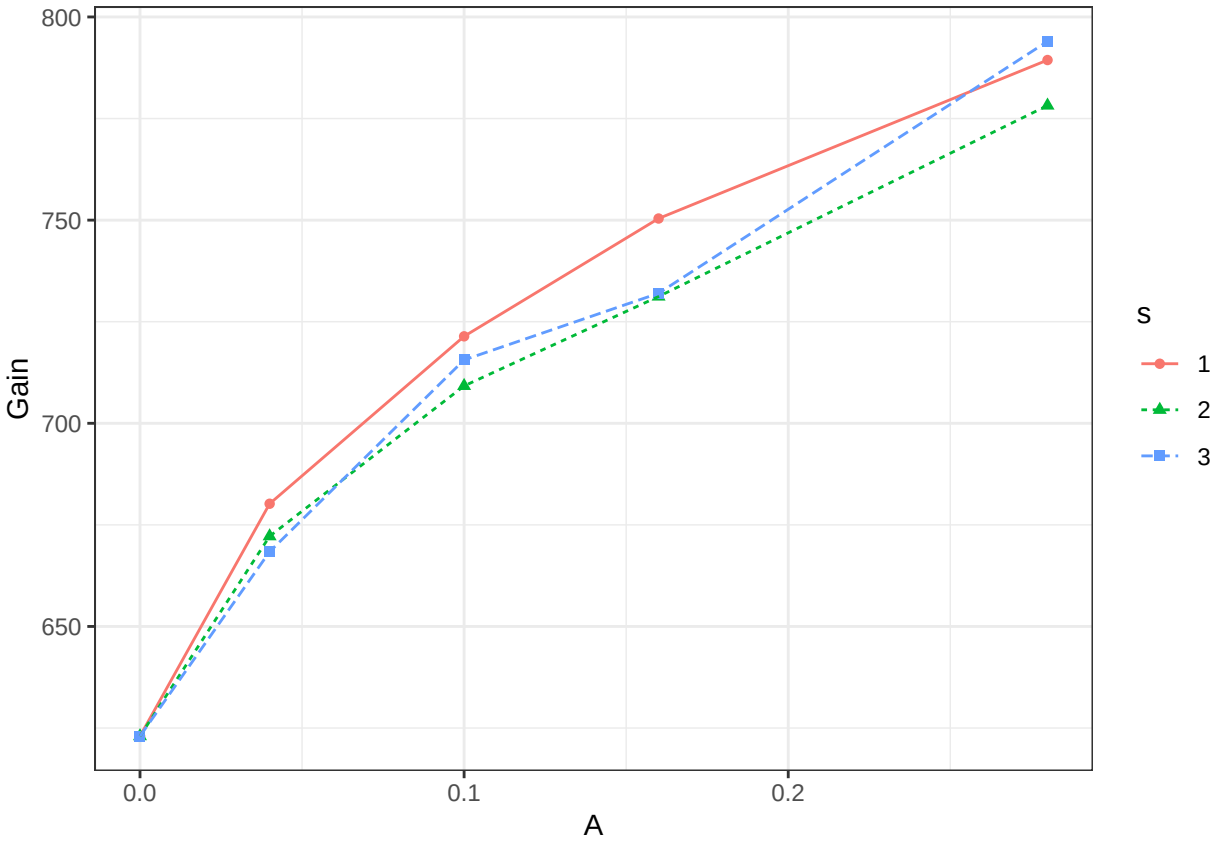
4 Predicting the Weather

```
ggplot(data = ftcollinssnow, mapping = aes(Early,Late))+
  geom_point(shape = 21,size =2.5)+
  ggtitle("Late~Early")+
  geom_smooth(method = "lm",linetype ="dashed",se = FALSE,
              col = "pink",size =1.5)+
  geom_hline(yintercept = mean(ftcollinssnow$Late),col ="steelblue",size =1.5)+
  theme_bw()
```



5 Turkey Growth

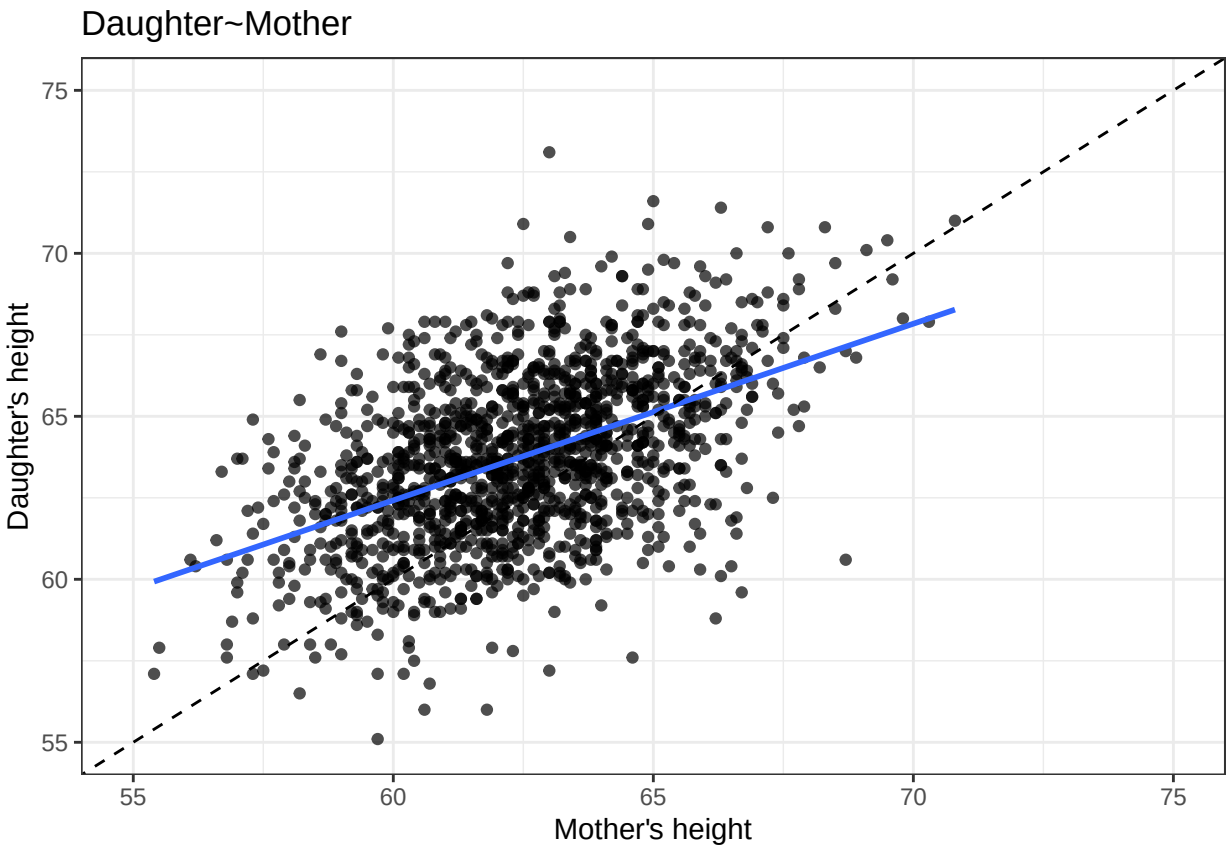
```
data <- rbind(turkey,data.frame(A =rep(0,3),Gain = rep(623,3),S = 1:3,m = rep(0,3),SD = rep(15,3))) %>%
  mutate(s = as.factor(S))


ggplot(data = data,mapping = aes(A,Gain,col= s))+
  geom_point(aes(shape = s),)+
  geom_line(aes(group = s,linetype = s))+
  theme_bw()
```



6 MEAN FUNCTION

$$E(dheight|mheight = x) = \beta_o + \beta_1 x$$

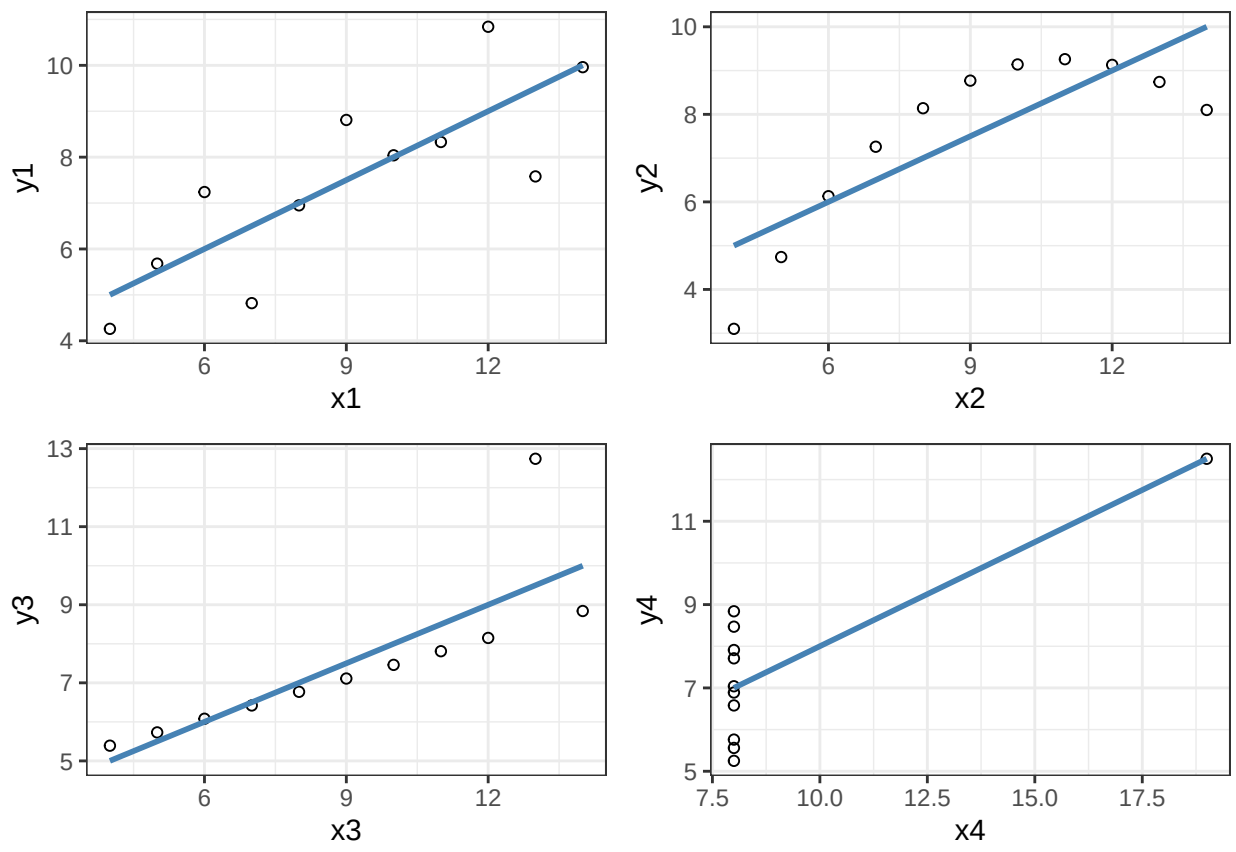
```
ggplot(data = Heights,mapping = aes(mheight,dheight))+
  geom_point(alpha = .7, col = "black")+
  ggtitle("Daughter~Mother")+
  xlab("Mother's height")+
  ylab("Daughter's height")+
  xlim(55,75)+
  ylim(55,75)+
  geom_smooth(method = "lm",se =FALSE)+
  geom_abline(slope = 1,intercept = 0,linetype = "dashed")+
  theme_bw()
```



7 SUMMARY GRAPH

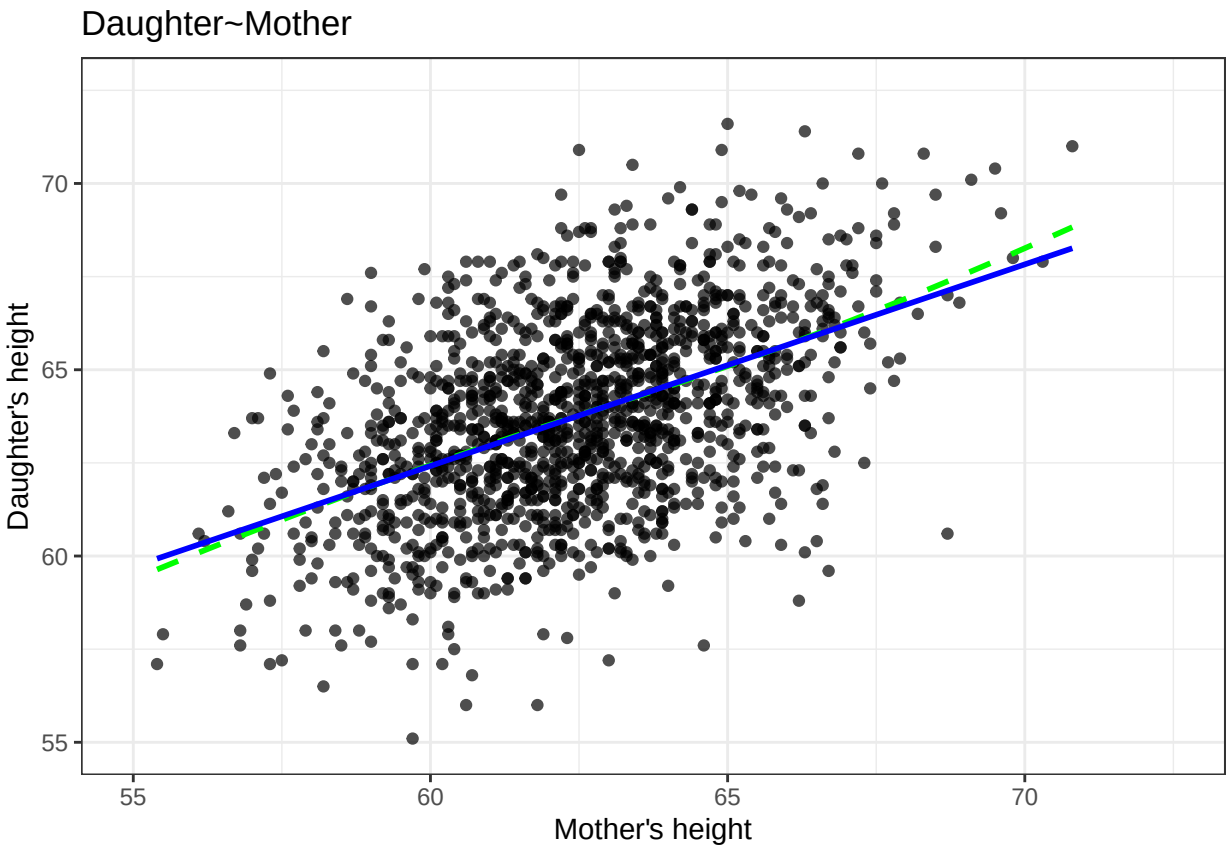
```
an1 <- ggplot(data = anscombe, mapping = aes(x1 ,y1))+
  geom_point(shape = 21,size =1.5)+
  geom_smooth(se = FALSE,method = "lm", col = "steelblue")+
  theme_bw()

an2 <- ggplot(data = anscombe, mapping = aes(x2 ,y2))+
  geom_point(shape = 21,size =1.5)+
  geom_smooth(se = FALSE,method = "lm", col = "steelblue")+
  theme_bw()

an3 <- ggplot(data = anscombe, mapping = aes(x3 ,y3))+
  geom_point(shape = 21,size =1.5)+
  geom_smooth(se = FALSE,method = "lm", col = "steelblue")+
  theme_bw()

an4 <- ggplot(data = anscombe, mapping = aes(x4 ,y4))+
  geom_point(shape = 21,size =1.5)+
  geom_smooth(se = FALSE,method = "lm", col = "steelblue")+
  theme_bw()
```

```
grid.arrange(an1,an2,an3,an4,ncol =2)
```



8 Smoothers for the Mean Function

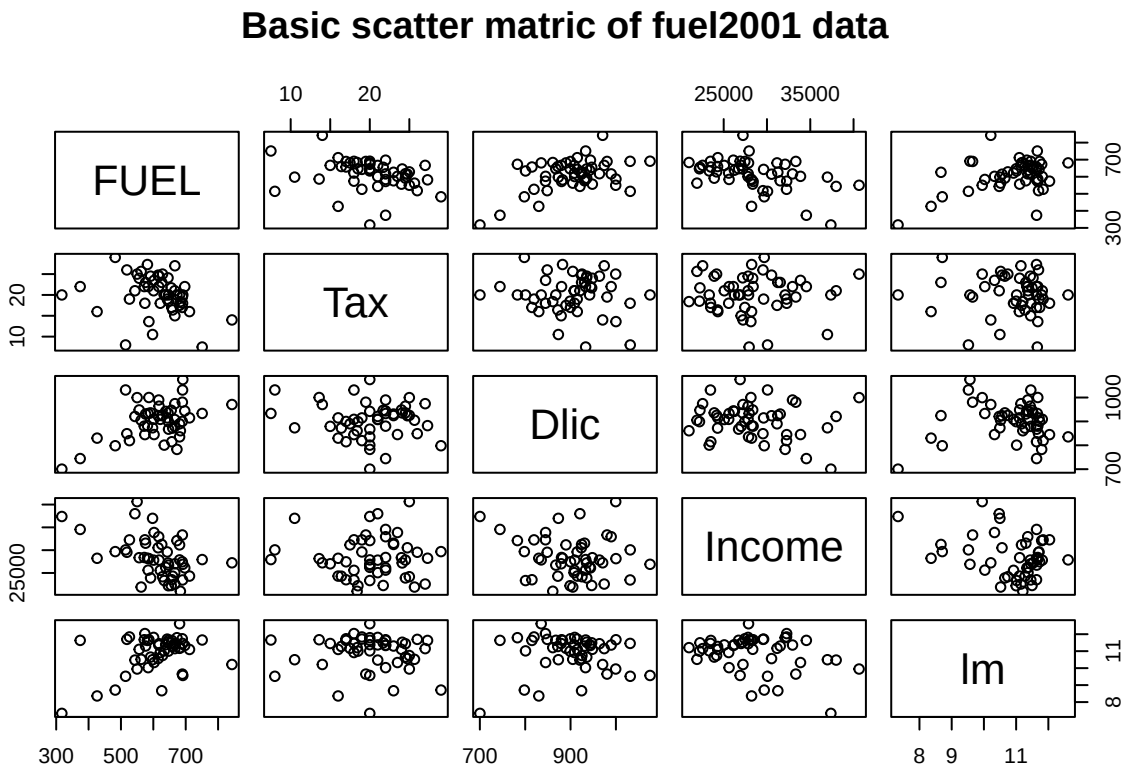
```
ggplot(data = Heights,mapping = aes(mheight,dheight))+
  geom_point(alpha = .7, col = "black")+
  ggtitle("Daughter~Mother")+
  xlab("Mother's height")+
  ylab("Daughter's height")+
  xlim(55,72.5)+
  ylim(55,72.5)+
  theme_bw()+
  geom_smooth(method = "loess",se = FALSE,col = "green",linetype = "dashed",
              span = 1.1  )+
  geom_smooth(method = "lm", se = FALSE, col = "blue")
```



9 SCATTERPLOT MATRICES

9.1 FUEL Consumption

```
f1<- as_tibble(fuel2001)

   f1 <- f1%>%
  mutate(FUEL = 1000*(FuelC/Pop),Dlic =1000*(Drivers/Pop),lm = log(Miles))%>%
     select(FUEL,Tax,Dlic,Income,lm)

pairs(f1[,1:5], main = "Basic scatter matric of fuel2001 data")
```



9.2 Correlation Matrix and Correlation coefficient(ρ):

```
ggpairs(f1[,1:5],
        title = "scatterplot matrix of fuel2001 data",
        upper = list(continuous = wrap("cor", size = 3)),
        lower = list(continuous = wrap("smooth",se =FALSE,method ="lm")))+
        theme_bw()
```

scatterplot matrix of fuel2001 data