CHAPTER 1 PROBLEMS

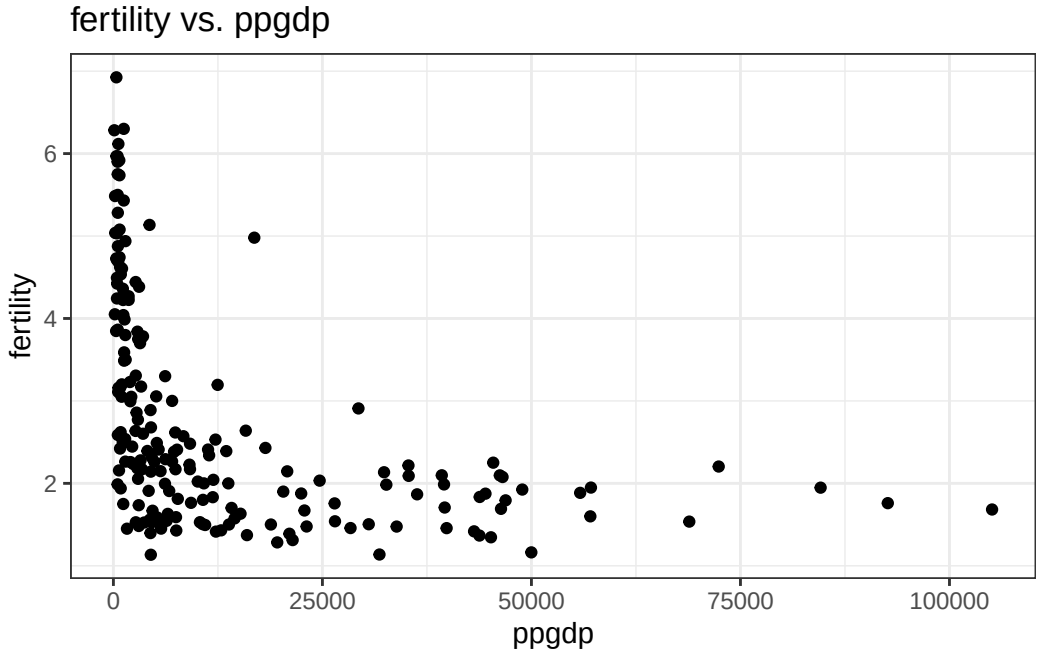
Andaleeb Hassan

1.1

The predictor is *ppgdp* and the response is $fertility$.

1.2

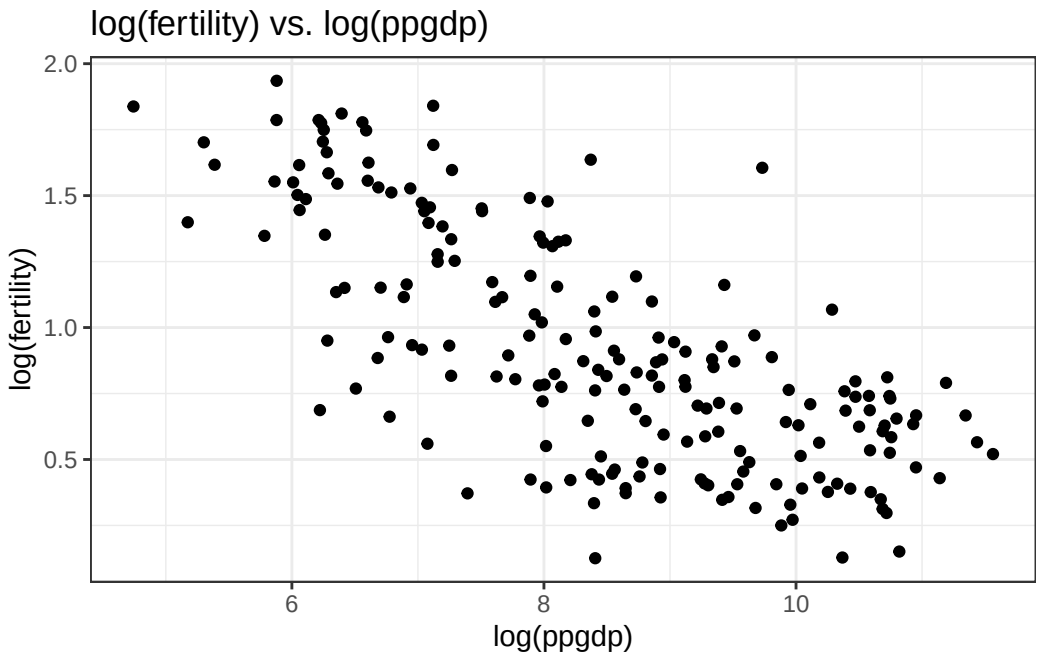
```
ggplot(UN11, mapping =aes(ppgdp,fertility))+
  geom_point()+
  ggtitle("fertility vs. ppgdp")+
  theme_bw()
```



We can't see a linear association from the scatterplot and thus a straight line mean function is not appropriate.

1.3

```
ggplot( UN11, mapping =aes( log(ppgdp),log(fertility)))+
  geom_point()+
  ggtitle("log(fertility) vs. log(ppgdp)")+
  theme_bw()
```



A simple linear regression model $E(\log(fertility)|\mathbf{log(ppgdp)}) = \beta_0 + \beta_1 \log(ppgdp)$ seems plausible.

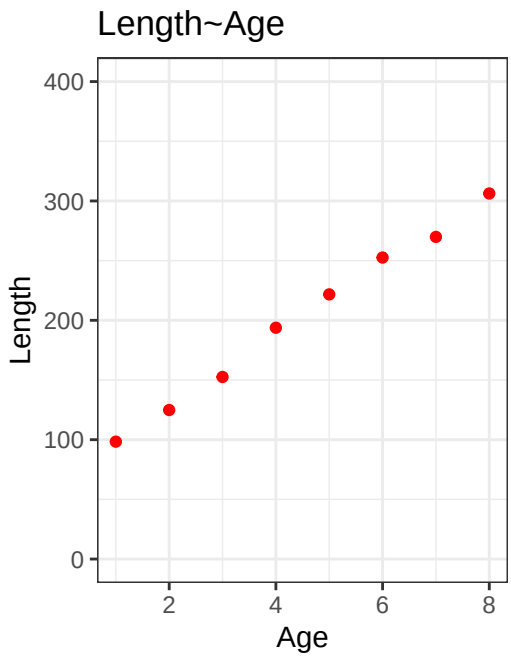
2

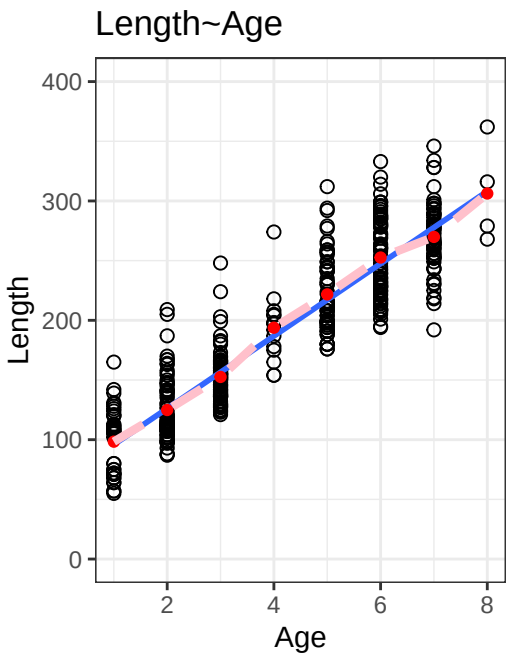
```
wblake%>%
  group_by(Age)%>%
  summarise(mean = mean(Length),
            var  =  var(Length))
```

```
# A tibble: 8 x 3
    Age  mean   var
  <int> <dbl> <dbl>
1     1  98.3  808.
2     2 125.   697.
3     3 153.   412.
4     4 194.   867.
5     5 222.   986.
6     6 253.  1105.
7     7 270.   869.
8     8 306.  1803.
```

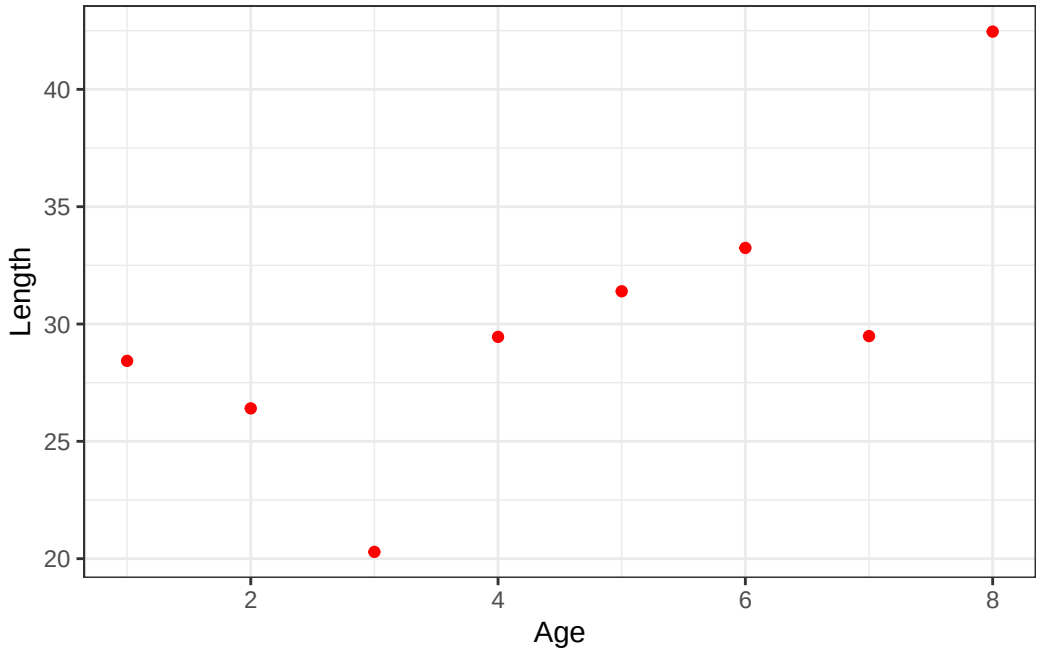
```
M1 <- ggplot( data = wblake,mapping = aes(Age,Length))+
  #geom_point( shape = 21, size =2)+
  ggtitle("Length~Age")+
  #geom_smooth(se = FALSE,method = "lm")+
  stat_summary(geom = "point",fun = "mean",col ="red")+
  ylim(0,400)+
  #stat_summary(geom =  "line",fun = "mean",col = "pink",linewidth = 1.5,
               #linetype = "dashed")+
  theme_bw()

M2 <- 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 = "mean",col ="red")+
  stat_summary(geom =  "line",fun = "mean",col = "pink",linewidth = 1.5,
               linetype = "dashed")+
  ylim(0,400)+
  theme_bw()

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

```
`geom_smooth()` using formula = 'y ~ x'
```






```
ggplot( data = wblake,mapping = aes(Age,Length))+
  stat_summary(geom = "point",fun = "sd",col ="red")+
  theme_bw()
```

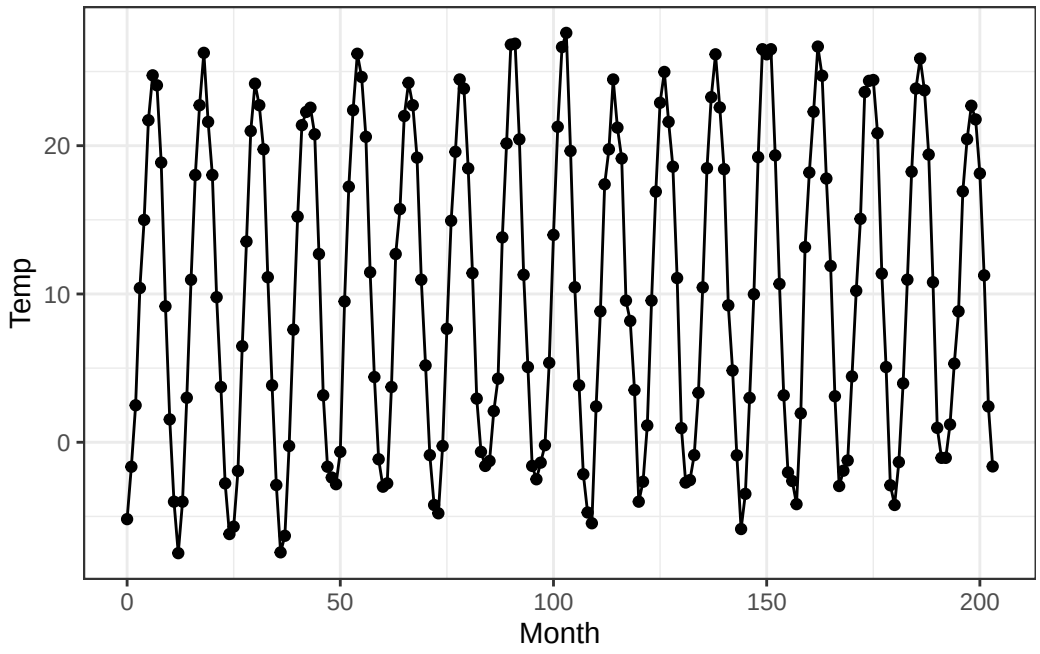


The plot is not a null plot.

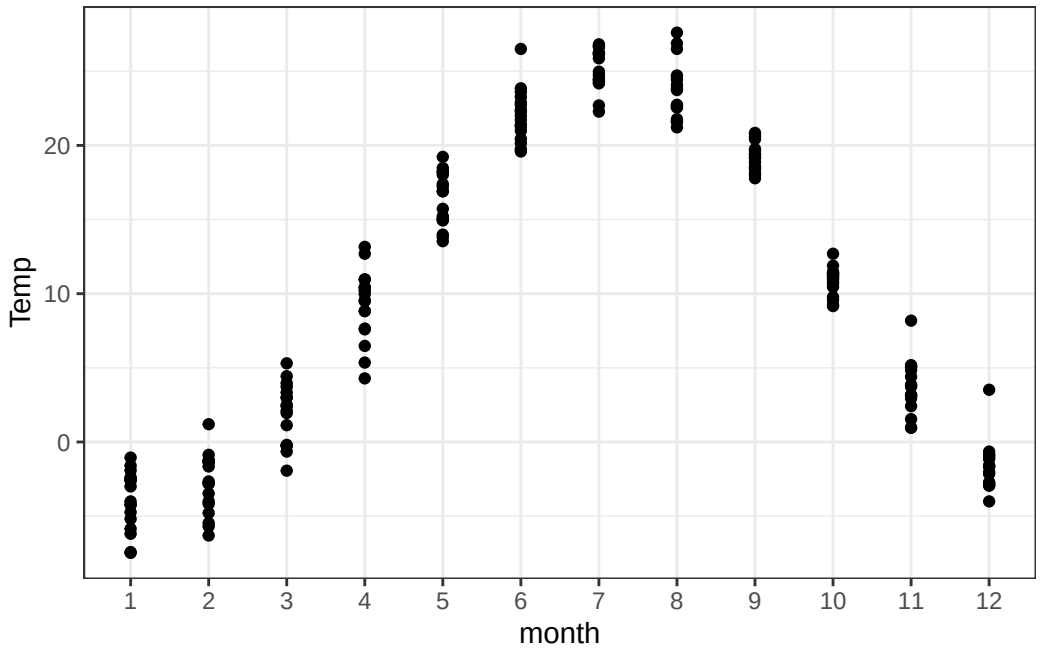
3.1

This is a time series data so simple linear regression wouldn't be adequate for modelling the data.

```
ggplot(Mitchell,mapping = aes(Month,Temp))+
  geom_point()+
  geom_line()+
  theme_bw()
```




```
data <- as_tibble(Mitchell) %>%
  mutate(month = as.factor((Month%%12)+1))


ggplot(data = data , mapping = aes(month,Temp))+
  geom_point()+
  theme_bw()
```



We can treat months as factors and fit the following regression model:

$$E(Temp|month) = \delta_0 + \delta_1 U_2 + \cdots + \delta_{11} U_{12}.$$

Where,

$$U_i = \begin{cases} 1 & \text{if, month} = \text{i} \\ 0 & \text{if, month} \neq i \end{cases} \qquad [i = 2, \cdots, 12]$$

The fitted model:

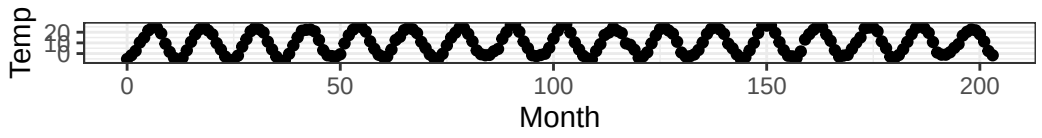
```
model <- lm(Temp~month,data)

broom::tidy(model)%>%
  select(1:2)%>%
  mutate( estimate = round(estimate,3))
```

```
# A tibble: 12 x 2
   term        estimate
   <chr>          <dbl>
 1 (Intercept)    -3.94
 2 month2          1.07
 3 month3          5.91
 4 month4         13.2
```

```
 5 month5          20.3
 6 month6          26.0
 7 month7          29.1
 8 month8          27.7
 9 month9          23.2
10 month10         14.8
11 month11          7.61
12 month12          2.27
```

3.2

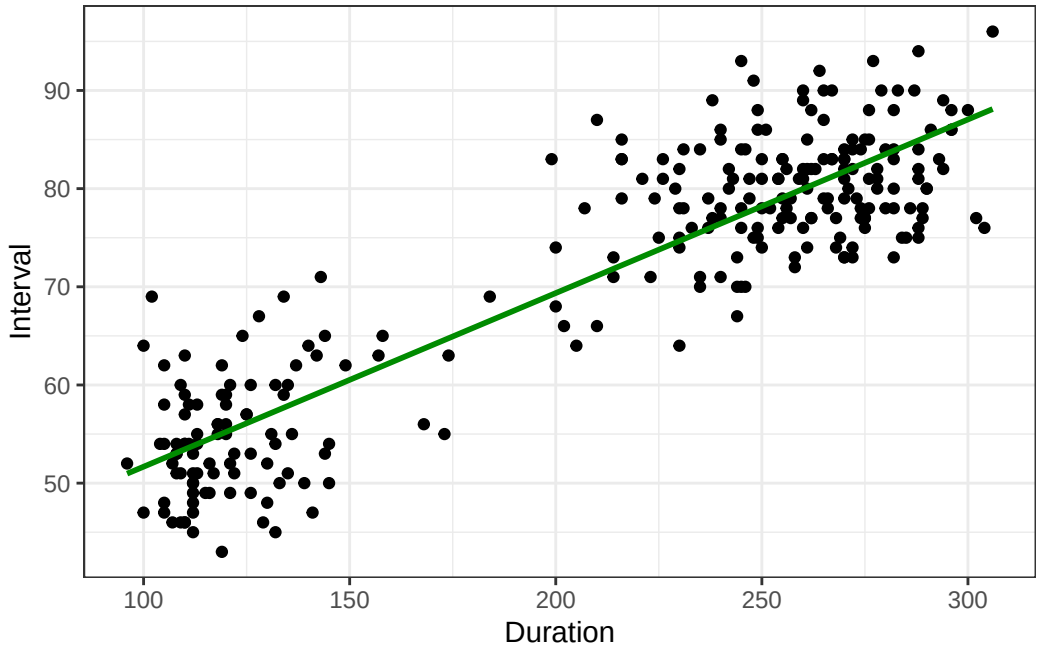
```
ggplot(Mitchell,mapping = aes(Month,Temp))+
  geom_point()+
  #geom_line()+
  coord_fixed(ratio =  1/4)+
  theme_bw()
```



4

```
ggplot( data = oldfaith, mapping = aes(Duration,Interval))+
  geom_point()+
  geom_smooth(method = "lm",se =FALSE, col = "green4")+
  theme_bw()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
model <- lm(Interval~Duration,oldfaith)
broom::tidy(model)%>%
  select(1:3)
```

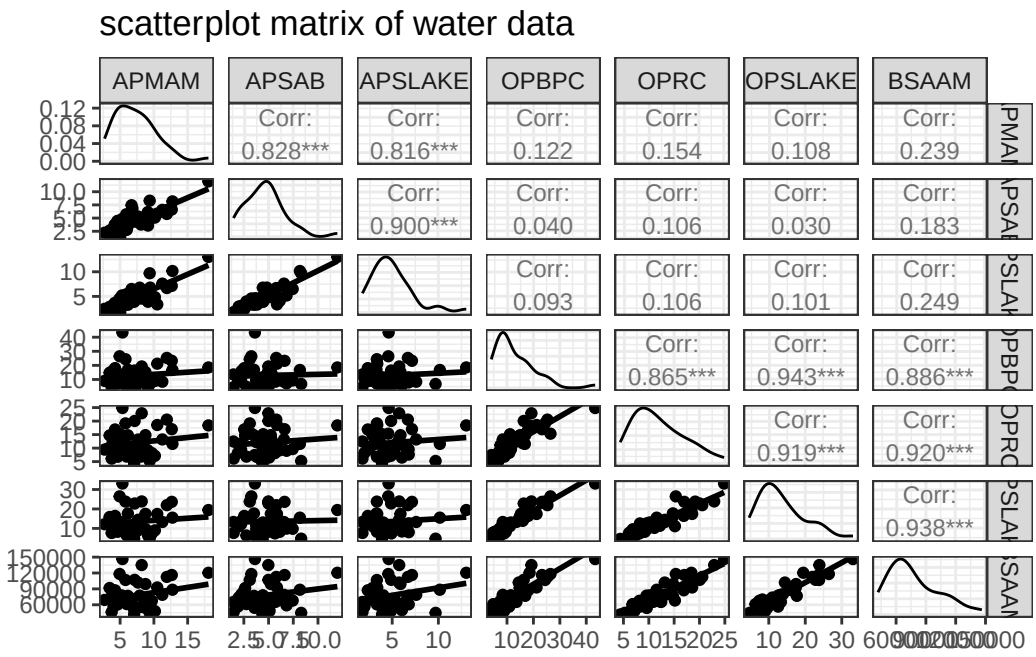
```
# A tibble: 2 x 3
  term        estimate std.error
  <chr>          <dbl>     <dbl>
1 (Intercept)   34.0     1.18
2 Duration       0.177   0.00535
```

The fitted linear regression model:

$$\hat{E}(Interval|Duration) = 33.99 + .18\ Duration\ .$$

5

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



6

```
data <- Rateprof %>%
  select(quality,clarity,helpfulness,easiness,raterInterest)
ggpairs(data,
```

```
        title = "scatterplot matrix of water data",
        upper = list(continuous = wrap("cor", size = 3)),
        lower = list(continuous = wrap("smooth",se =FALSE,method ="lm")))+
  theme_bw()
```