

مجموعه تمرین های چهارم

۱. یک معلم، دانش آموزان خود را با مقیاس ۱ تا ۵ ارزیابی کرده است. فرض کنید پاسخ به سه سوال اول در این جدول داده شده است:

Student	Ques. 1	Ques. 2	Ques. 3
1	3	5	1
2	3	2	3
3	3	5	1
4	4	5	1
5	3	2	1
6	4	2	3
7	3	5	1
8	4	5	1
9	3	4	1
10	4	2	1

با استفاده از یکی از دستورهای `c()`, `scan()`, `read.table` or `data.entry()` داده ها را وارد کنید.

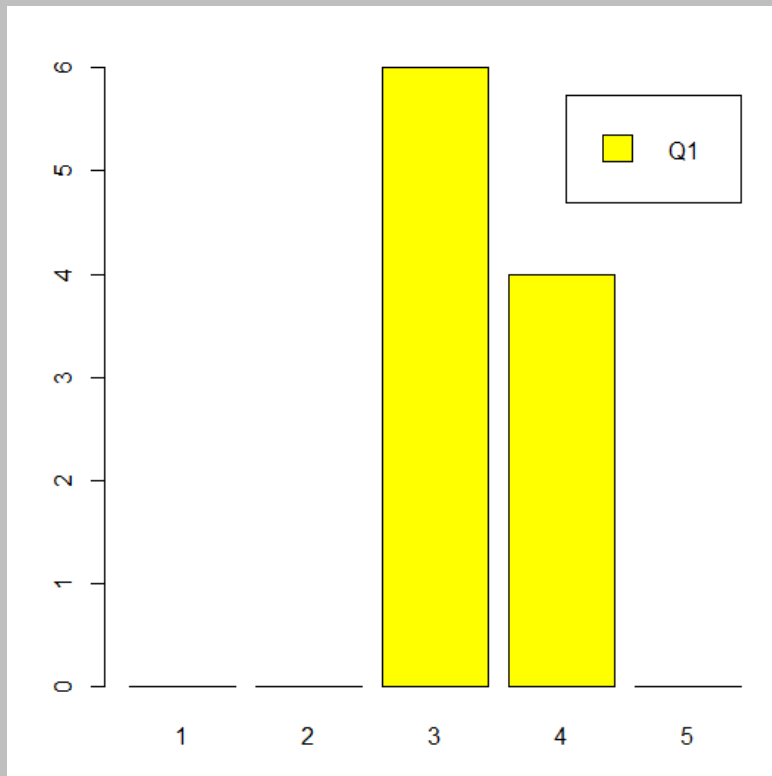
۱. جدولی از نتایج سوال ۱ و سوال ۲ را جداگانه تهیه کنید.

۲. جدول احتمالی (contingency table) سوالات ۱ و ۲ را ایجاد کنید.

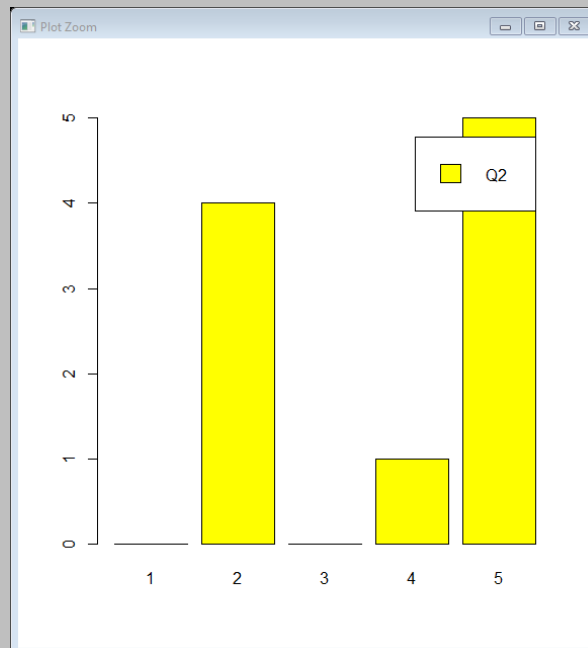
۳. یک `barplot` از سوالات ۲ و ۳ ایجاد کنید.

۴. برای هر ۳ سوال یک `barplot` ایجاد کنید و ۳ نمودار را در یک شکل کنار هم نمایش دهید.

```
1 a <- read.table(textConnection("1 3 5 1
2 2 3 2 3
3 3 3 5 1
4 4 4 5 1
5 5 3 2 1
6 6 4 2 3
7 7 3 5 1
8 8 4 5 1
9 9 3 4 1
0 10 4 2 1"))
1 a <- t(a[-1])
2 rownames(a) <- c('Q1', 'Q2', 'Q3')
3 q1 <- factor(rep('Q1', 10))
4 r1 <- factor(a[1,], levels=c(1,2,3,4,5))
5 table.q1 <- table(q1, r1)
6 barplot(table.q1,col='yellow',legend.text=T)
```



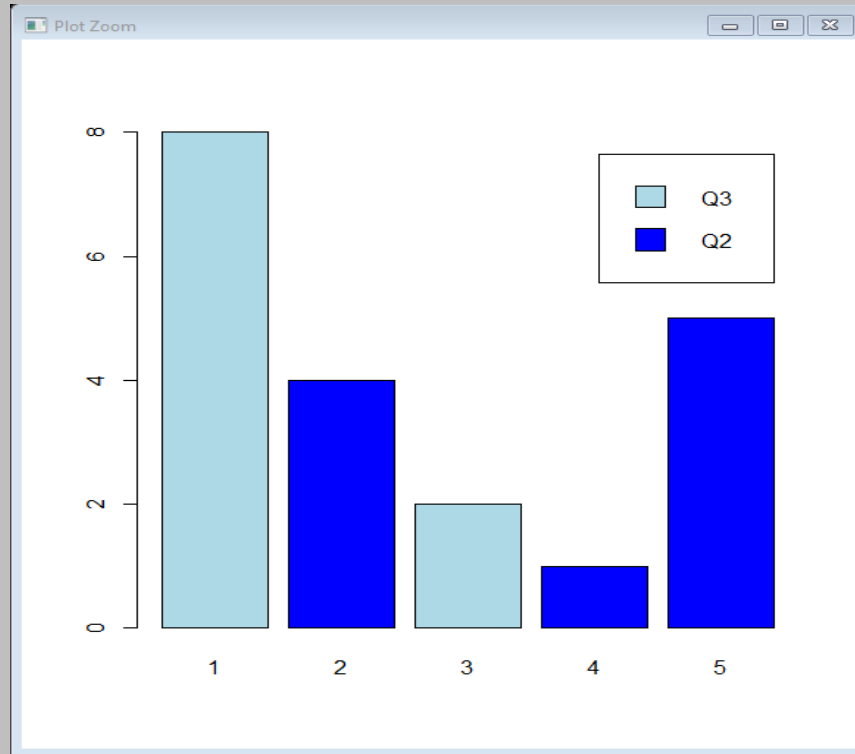
```
q2 <- factor(rep('Q2', 10))  
r2 <- factor(a[2,], levels=c(1,2,3,4,5))  
table.q2 <- table(q2, r2)  
barplot(table.q2, col='yellow', legend.text=T)  
"
```



```

q1_2 <- factor(rep(c('Q1', 'Q2'), each=10))
r1_2 <- factor(append(a[1,], a[2,]), levels=c(1,2,3,4,5))
table(q1_2, r1_2)
#1-3
q2_3 <- factor(rep(c('Q2', 'Q3'), each=10))
r2_3 <- factor(append(a[2,], a[3,]), levels=c(1,2,3,4,5))
table.q2_3 = table(q2_3, r2_3)
barplot(table.q2_3,col=c('blue', 'lightblue'),legend=T)

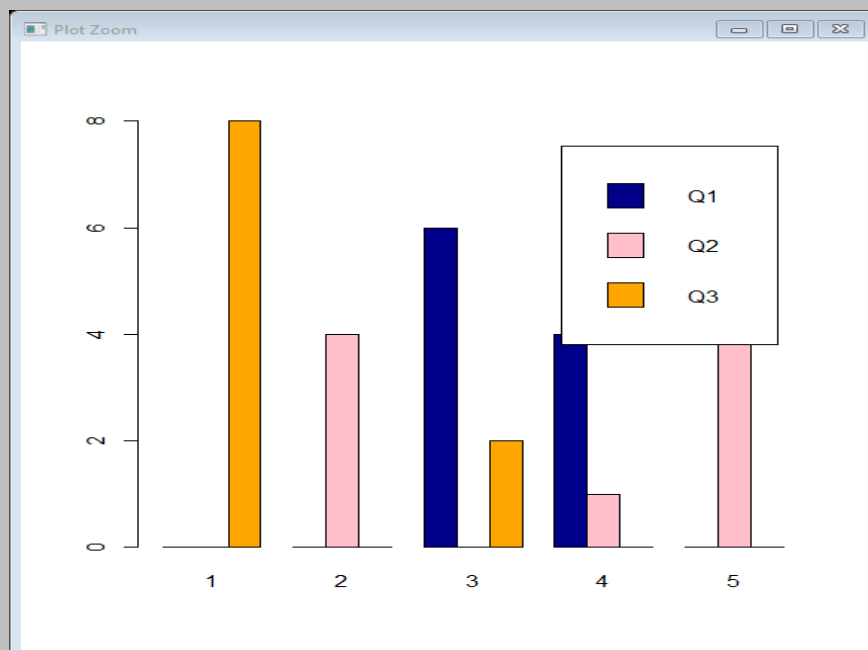
```



```

q.all <- factor(rep(c('Q1', 'Q2', 'Q3'), each=10))
r.all <- factor(append(append(a[1,], a[2,]), a[3,]), levels=c(1,2,3,4,5))
table.all <- table(q.all, r.all)
barplot(table.all,col=c('darkblue', 'pink', 'orange'),legend=T,beside=T)

```



۲. در پکیج MASS مجموعه ای از داده های UScereal وجود دارد که حاوی اطلاعاتی درباره غلات صبحانه ایی است.
. مجموعه داده ها را به صورت زیر قرار دهید:

```
> library('MASS')  
> data('UScereal')  
> attach(UScereal)  
> names(UScereal) # to see the names
```

اکنون، روابط زیر را بررسی کنید و نظرات خود را در مورد آنچه که مشاهده می کنید، بیان کنید. شما می توانید از tables, scatterplots, barplots, و غیره برای بررسی روابط استفاده کنید.

۱. رابطه بین سازنده و قفسه (manufacturer and shelf)

۲. ارتباط چربی و ویتامین ها (fat and vitamins)

۳. رابطه چربی و قفسه (fat and shelf)

۴. رابطه بین کربوهیدرات ها و قندها (carbohydrates and sugars)

۵. رابطه بین فیبر و سازنده (fibre and manufacturer)

۶. رابطه بین سدیم و قندها (sodium and sugars)

آیا روابط دیگری وجود دارد که می توانید پیش بینی و تحقیق کنید؟

```
> ##2
```

```
> library('MASS')
```

```
> data('UScereal')
```

```
> attach(UScereal)
```

The following objects are masked from UScereal (pos = 3):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

The following objects are masked from UScereal (pos = 4):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

The following object is masked from package:UsingR:

fat

The following objects are masked from UScereal (pos = 12):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

The following objects are masked from UScereal (pos = 13):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

The following objects are masked from UScereal (pos = 14):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

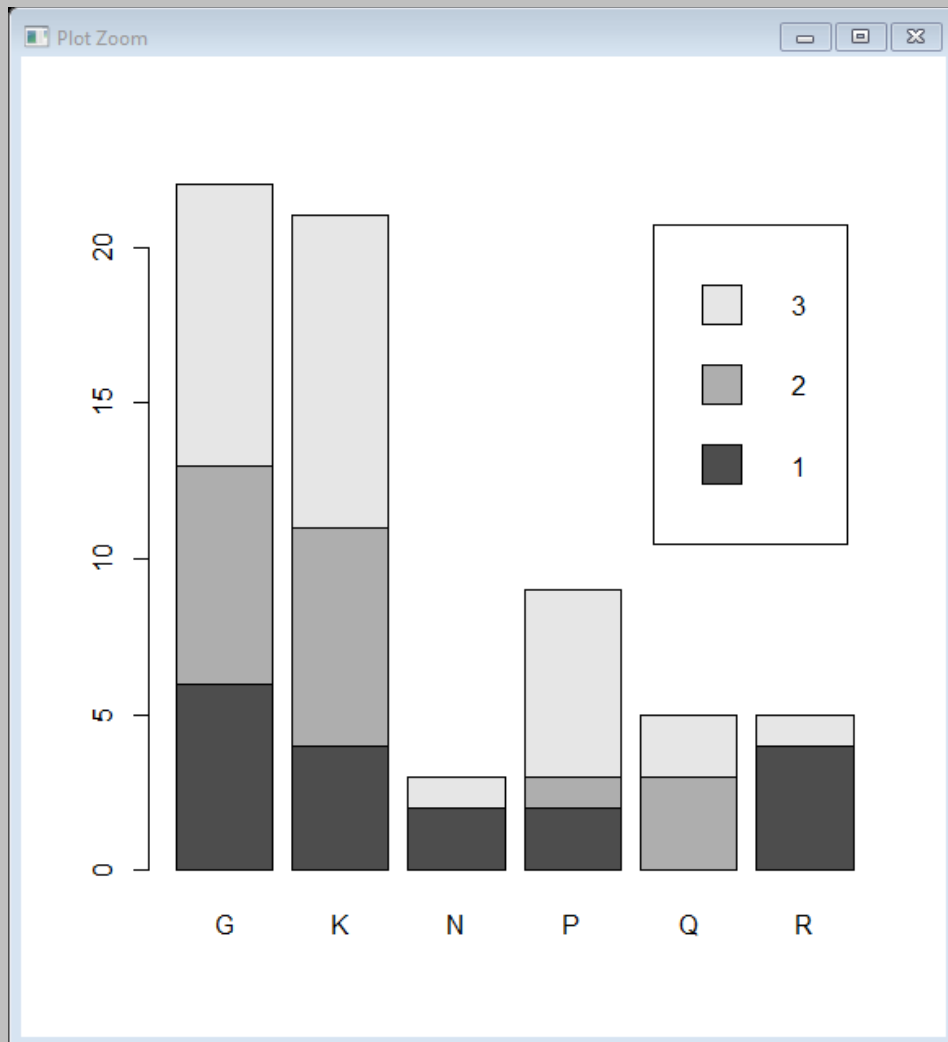
The following objects are masked from UScereal (pos = 15):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

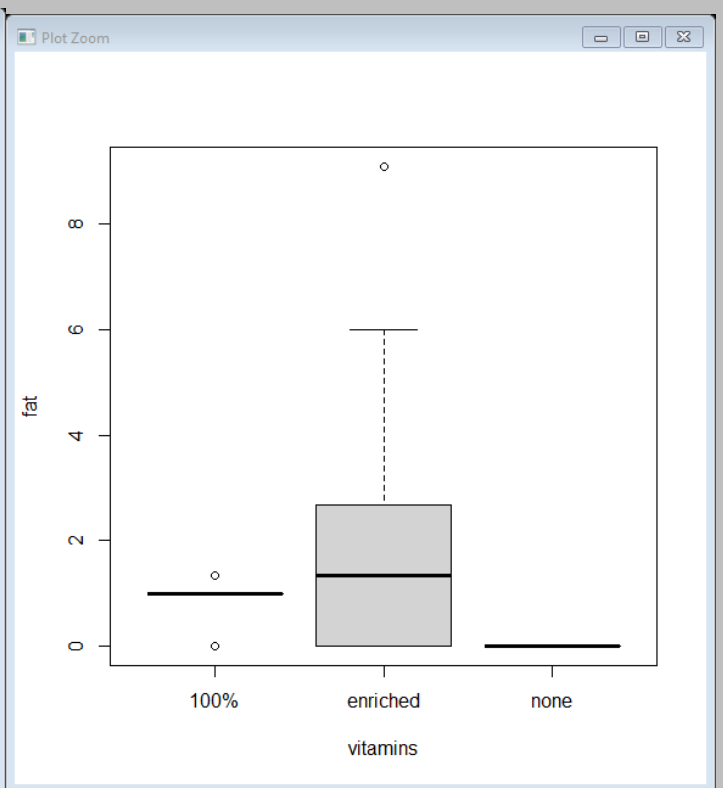
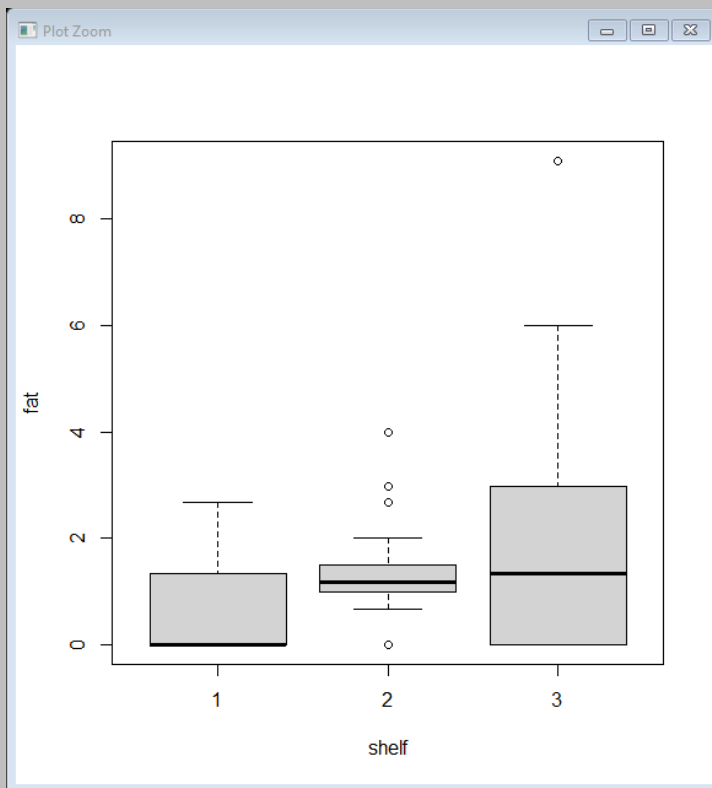
The following objects are masked from UScereal (pos = 16):

calories, carbo, fat, fibre, mfr, potassium, protein, shelf, sodium, sugars, vitamins

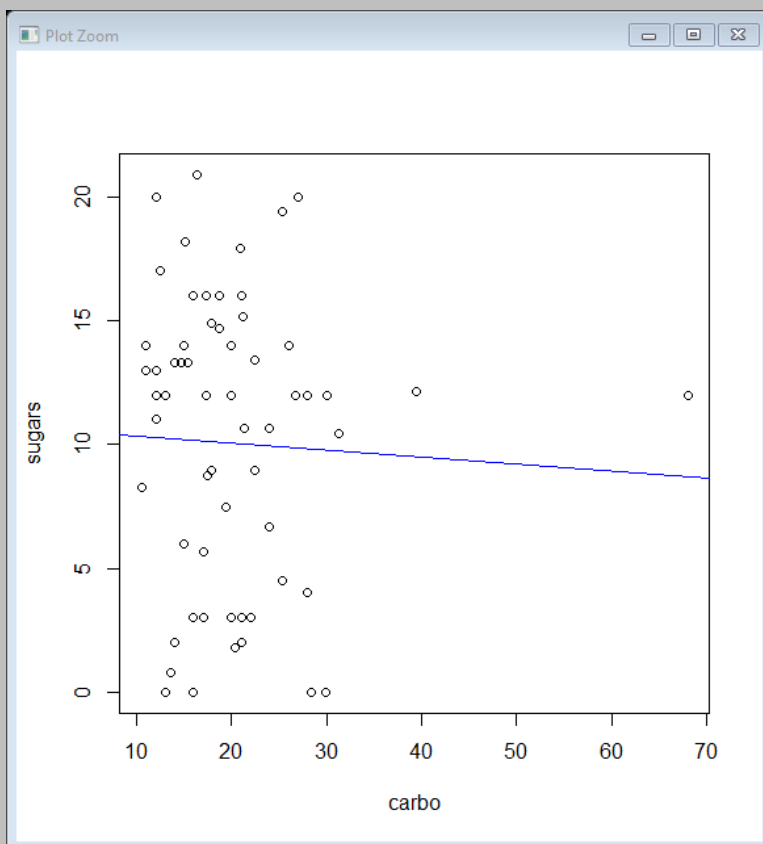
```
> names(UScereal)
[1] "mfr"      "calories" "protein"  "fat"      "sodium"   "fibre"
[7] "carbo"    "sugars"   "shelf"    "potassium" "vitamins"
> c <- UScereal
> barplot(table(c$shelf, c$mfr), legend=T)
```



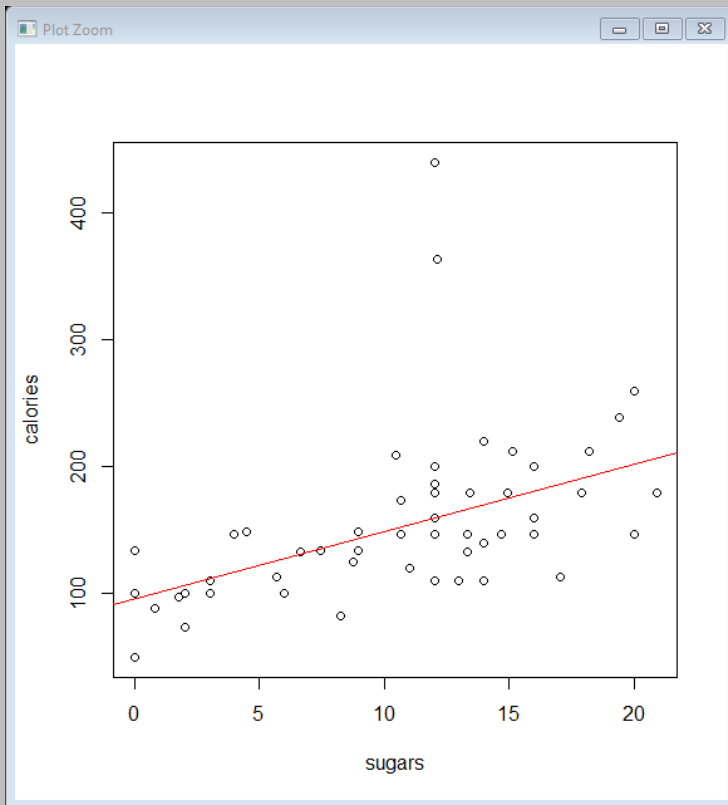
```
boxplot(fat ~ vitamins)
boxplot(fat ~ shelf)
```



```
plot(carbo, sugars)
lm <- lm(sugars ~ carbo)
abline(lm, col="blue")
```



```
plot(sugars, calories)
lm <- lm(calories ~ sugars)
abline(lm, col="red")
```



۳. مجموعه داده های **mammals** حاوی داده های مربوط به وزن بدن در مقابل وزن مغز پستانداران است. برای پیدا کردن ضریب همبستگی پیرسون و اسپیرمن ، از **COR** استفاده کنید. آیا آنها مشابه هستند؟ داده ها را با استفاده از دستور **plot** ترسیم کنید و ببینید. آیا انتظار دارید که آنها مشابه باشند؟ شما باید از این **plot** ها ناراضی باشید. بعد، لگاریتم (**log**) هر متغیر را **plot** کنید و ببینید که آیا تفاوت دارند.

```

> ##3
> data(package='MASS', mammals)
> attach(mammals)
The following objects are masked from mammals (pos = 3):

    body, brain

> summary(mammals)
      body      brain
Min.   : 0.005   Min.   : 0.14
1st Qu.: 0.600   1st Qu.: 4.25
Median : 3.342   Median : 17.25
Mean   : 198.790 Mean   : 283.13
3rd Qu.: 48.202  3rd Qu.: 166.00
Max.   :6654.000 Max.   :5712.00
> cor(body, brain, method='pearson')
[1] 0.9341638
> cor(body, brain, method='spearman')
[1] 0.9534986
> plot(body, brain)
> s <- lm(brain ~ body)
> abline(s, col="red")
> summary(s)

Call:
lm(formula = brain ~ body)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-810.07  -88.52  -79.64  -13.02  2050.33

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 91.00440   43.55258    2.09   0.0409 *
body         0.96650    0.04766   20.28  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 334.7 on 60 degrees of freedom
Multiple R-squared:  0.8727,    Adjusted R-squared:  0.8705
F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16

> plot(log1p(body), log1p(brain))
> e <- lm(log1p(brain) ~ log1p(body))
> abline(e, col="red")
> summary(e)

```



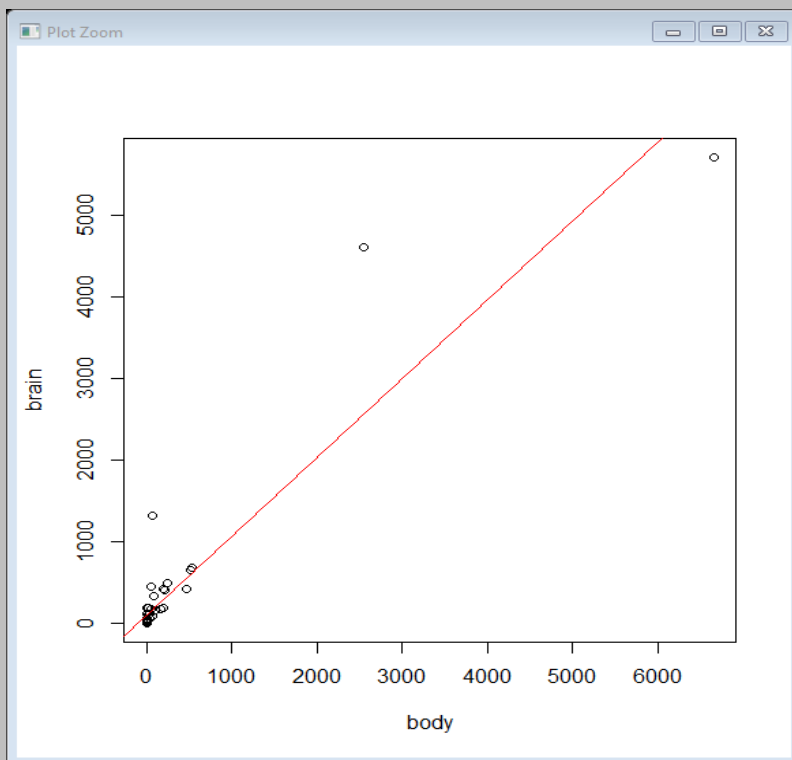
```
Call:
lm(formula = log1p(brain) ~ log1p(body))

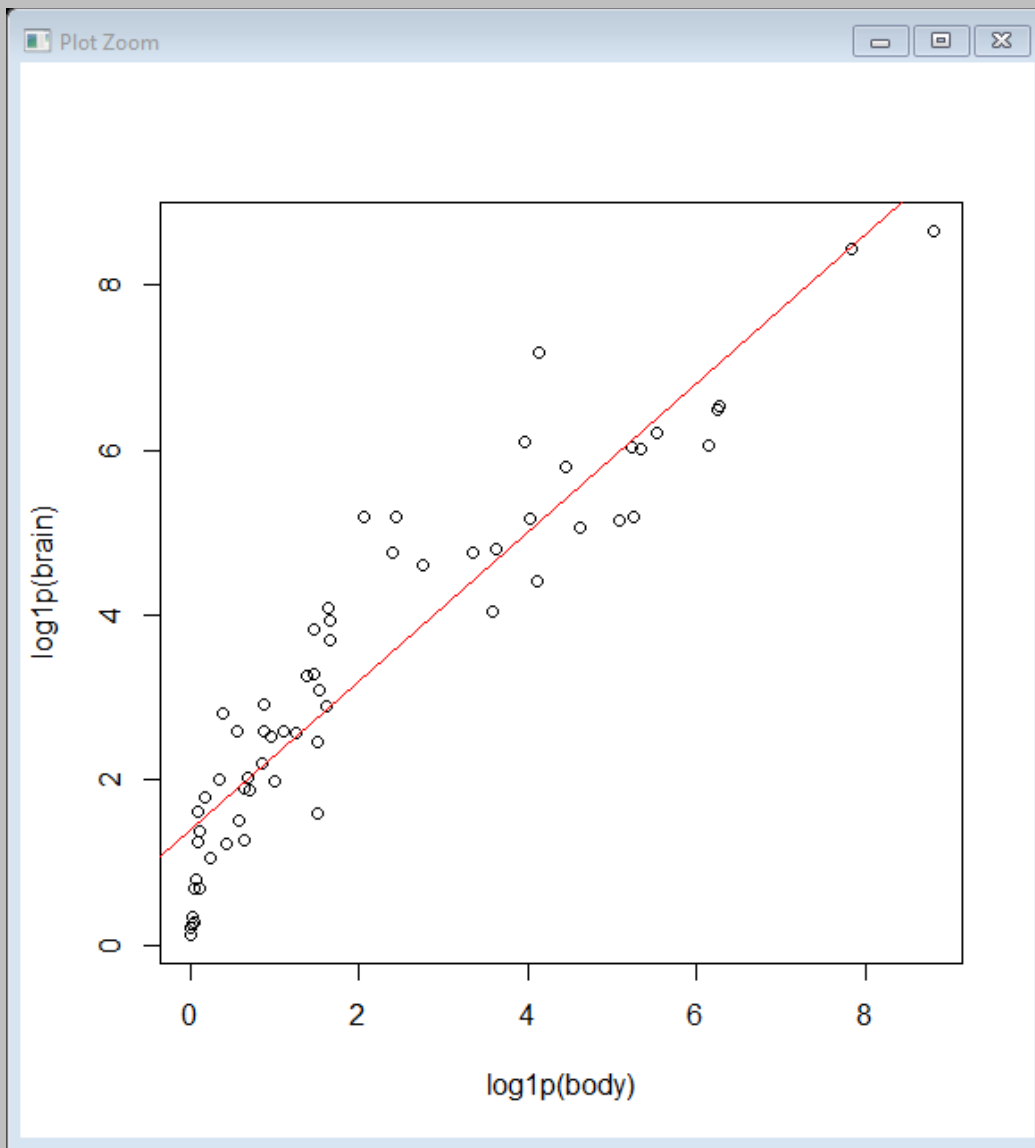
Residuals:
    Min       1Q   Median       3Q      Max
-1.27379 -0.57842 -0.04617  0.40597  2.05871

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.40033     0.14208   9.856 3.69e-14 ***
log1p(body)   0.89959     0.04578  19.652 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7927 on 60 degrees of freedom
Multiple R-squared:  0.8655,    Adjusted R-squared:  0.8633
F-statistic: 386.2 on 1 and 60 DF,  p-value: < 2.2e-16

> detach(mammals)
```





۴. برای داده ها قیمت مسکن، داده های **homedata** ، رابطه بین ارزش ارزیابی قدیمی (old) و جدید (new) را بررسی کنید. متغیر old را به عنوان متغیر پیشگو یا همان X ، متغیر جدید (y) را متغیر پاسخ نظر بگیرید. آیا داده ها یک رابطه خطی را نشان می دهند؟ آیا داده پرت وجود دارد یا خیر؟ چه چیزی ممکن است باعث ایجاد داده پرت شده باشد؟ پیش بینی ارزش ارزیابی جدید برای یک خانه ۷۵,۰۰۰ دلاری در سال ۱۹۷۰ چیست؟

```
> #4
> data(package='UsingR', homedata)
> attach(homedata)
The following objects are masked from homedata (pos = 3):
```

y1970, y2000

```
> summary(homedata)
      y1970      y2000
Min.   :    0   Min.   :   7400
1st Qu.: 57000  1st Qu.: 161400
Median : 68900  Median : 251700
Mean   : 70821  Mean   : 268370
3rd Qu.: 80500  3rd Qu.: 335600
Max.   :297200  Max.   :1182800
> plot(y1970, y2000)
> d <- lm(y2000 ~ y1970)
> abline(d, col="orange")
> summary(d)
```

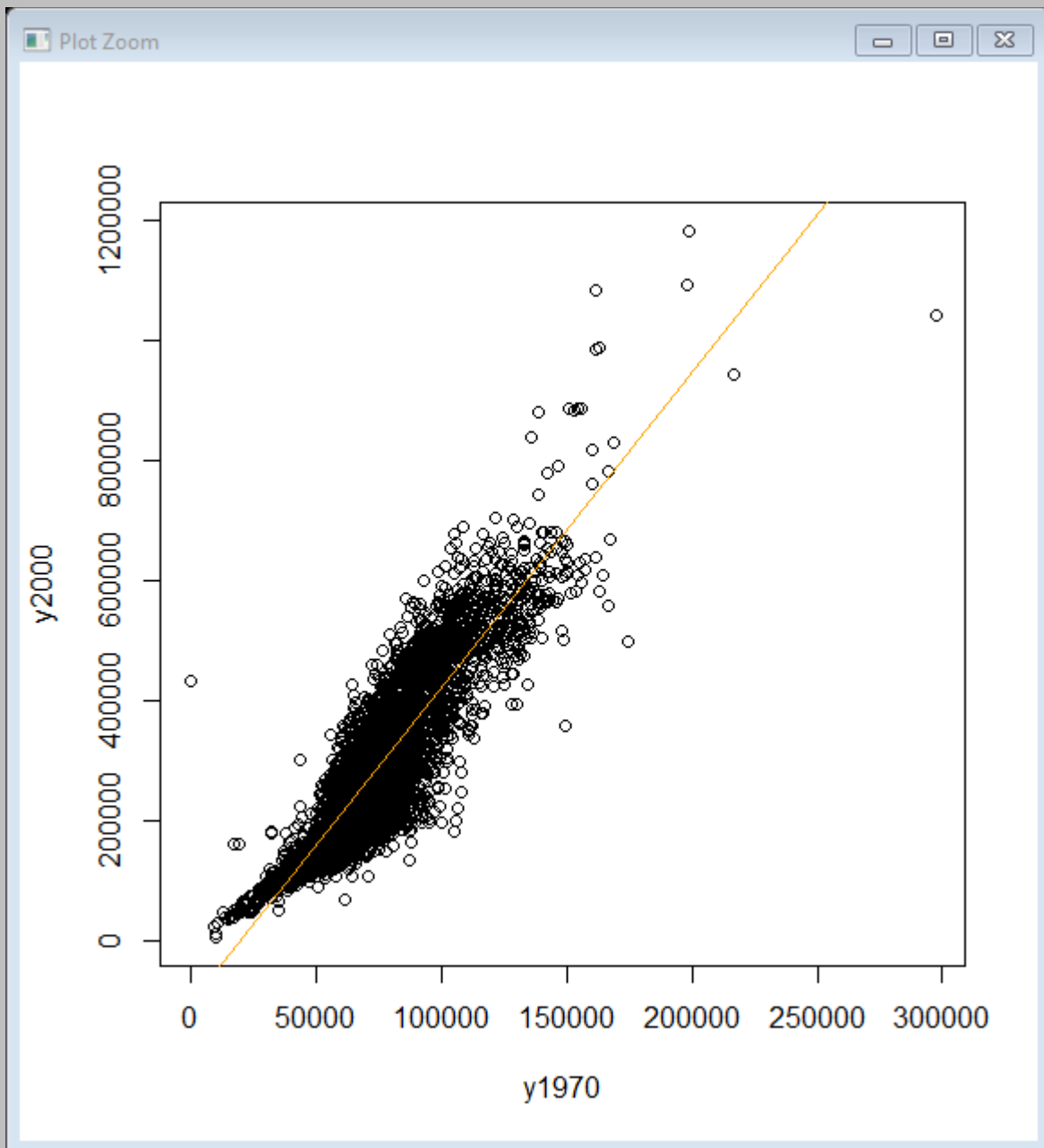
```
Call:
lm(formula = y2000 ~ y1970)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-416665  -36308      809   34372  536605
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.040e+05  2.337e+03  -44.51  <2e-16 ***
y1970        5.258e+00  3.147e-02  167.07  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 58000 on 6839 degrees of freedom
Multiple R-squared:  0.8032,    Adjusted R-squared:  0.8032
F-statistic: 2.791e+04 on 1 and 6839 DF,  p-value: < 2.2e-16
```

```
> a <- coef(d)[[1]]
> a
[1] -104004.7
> p <- coef(d)[[2]]
> p
[1] 5.257972
> predicted.price <- 75000*p + a
> predicted.price
[1] 290343.2
> detach(homedata)
```



۵. اطلاعات داده **babies** را ضمیمه کنید:

```
> library("Simple")  
> data("babies")  
> attach(babies)
```

این مجموعه داده شامل اطلاعات زیادی در مورد نوزادان و مادران آنها برای ۱۲۳۶ مشاهدات است. ضریب همبستگی (هر دو پیرسون و اسپیرمن) بین سن و وزن را پیدا کنید.
برای رابطه بین قد و وزن تکرار کنید.
نمودار پراکندگی هر یک از جفت ها را رسم کنید و ببینید آیا پاسخ شما منطقی است.

```
> ##5
> data(package="UsingR", babies)
> attach(babies)
The following objects are masked from babies (pos = 3):

  age, dage, date, ded, dht, drace, dwt, ed, gestation, ht, id, inc,
  marital, number, outcome, parity, plurality, race, sex, smoke, time, wt,
  wt1
```

```
> summary(babies)
```

id		plurality		outcome		date		gestation		sex	
Min.	: 15	Min.	:5	Min.	:1	Min.	:1350	Min.	:148.0	Min.	:1
1st Qu.	:5286	1st Qu.	:5	1st Qu.	:1	1st Qu.	:1444	1st Qu.	:272.0	1st Qu.	:1
Median	:6730	Median	:5	Median	:1	Median	:1540	Median	:280.0	Median	:1
Mean	:6001	Mean	:5	Mean	:1	Mean	:1536	Mean	:286.9	Mean	:1
3rd Qu.	:7583	3rd Qu.	:5	3rd Qu.	:1	3rd Qu.	:1627	3rd Qu.	:288.0	3rd Qu.	:1
Max.	:9263	Max.	:5	Max.	:1	Max.	:1714	Max.	:999.0	Max.	:1

wt		parity		race		age		ed	
Min.	: 55.0	Min.	: 0.000	Min.	: 0.000	Min.	:15.00	Min.	:0.000
1st Qu.	:108.8	1st Qu.	: 0.000	1st Qu.	: 0.000	1st Qu.	:23.00	1st Qu.	:2.000
Median	:120.0	Median	: 1.000	Median	: 3.000	Median	:26.00	Median	:2.000
Mean	:119.6	Mean	: 1.932	Mean	: 3.206	Mean	:27.37	Mean	:2.922
3rd Qu.	:131.0	3rd Qu.	: 3.000	3rd Qu.	: 7.000	3rd Qu.	:31.00	3rd Qu.	:4.000
Max.	:176.0	Max.	:13.000	Max.	:99.000	Max.	:99.00	Max.	:9.000

ht		wt1		drace		dage		ded	
Min.	:53.00	Min.	: 87	Min.	: 0.000	Min.	:18.00	Min.	:0.000
1st Qu.	:62.00	1st Qu.	:115	1st Qu.	: 0.000	1st Qu.	:25.00	1st Qu.	:2.000
Median	:64.00	Median	:126	Median	: 3.000	Median	:29.00	Median	:4.000
Mean	:64.67	Mean	:154	Mean	: 3.665	Mean	:30.74	Mean	:3.189
3rd Qu.	:66.00	3rd Qu.	:140	3rd Qu.	: 7.000	3rd Qu.	:35.00	3rd Qu.	:5.000
Max.	:99.00	Max.	:999	Max.	:99.000	Max.	:99.00	Max.	:9.000

dht		dwt		marital		inc		smoke	
Min.	:60.00	Min.	:110.0	Min.	:0.000	Min.	: 0.00	Min.	:0.0000
1st Qu.	:70.00	1st Qu.	:165.0	1st Qu.	:1.000	1st Qu.	: 2.00	1st Qu.	:0.0000
Median	:73.00	Median	:190.0	Median	:1.000	Median	: 4.00	Median	:1.0000
Mean	:81.67	Mean	:505.4	Mean	:1.038	Mean	:13.16	Mean	:0.8681
3rd Qu.	:99.00	3rd Qu.	:999.0	3rd Qu.	:1.000	3rd Qu.	: 7.00	3rd Qu.	:1.0000
Max.	:99.00	Max.	:999.0	Max.	:5.000	Max.	:98.00	Max.	:9.0000

time		number	
Min.	: 0.000	Min.	: 0.000
1st Qu.	: 0.000	1st Qu.	: 0.000
Median	: 1.000	Median	: 1.000
Mean	: 1.748	Mean	: 2.604
3rd Qu.	: 1.000	3rd Qu.	: 3.000
Max.	:99.000	Max.	:98.000

```
> cor(age, wt1, method="pearson")
[1] 0.06273172
> cor(age, wt1, method="spearman")
[1] 0.1453316
> babe <- lm(wt1 ~ age)
> plot(age, wt1)
> abline(babe, col="red")
```

```

> plot(babe, which=1)
> plot(babe, which=2)
> summary(babe)

Call:
lm(formula = wt1 ~ age)

Residuals:
    Min       1Q   Median       3Q      Max
-118.88  -39.13  -27.57  -12.13   857.05

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  114.6541    18.2974   6.266  5.1e-10 ***
age           1.4366     0.6506   2.208  0.0274 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 147.6 on 1234 degrees of freedom
Multiple R-squared:  0.003935, Adjusted R-squared:  0.003128
F-statistic: 4.875 on 1 and 1234 DF,  p-value: 0.02743

> cor(ht, wt1, method="pearson")
[1] 0.6010033
> cor(ht, wt1, method="spearman")
[1] 0.5036179
> babe <- lm(wt1 ~ ht)

```

```
> babe <- lm(wt1 ~ ht)
> plot(ht, wt1)
> abline(babe, col="red")
> plot(babe, which=1)
> plot(babe, which=2)
> summary(babe)
```

Call:

```
lm(formula = wt1 ~ ht)
```

Residuals:

Min	1Q	Median	3Q	Max
-614.89	-40.28	-15.77	13.18	940.80

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-938.4516	41.4926	-22.62	<2e-16 ***
ht	16.8924	0.6395	26.41	<2e-16 ***

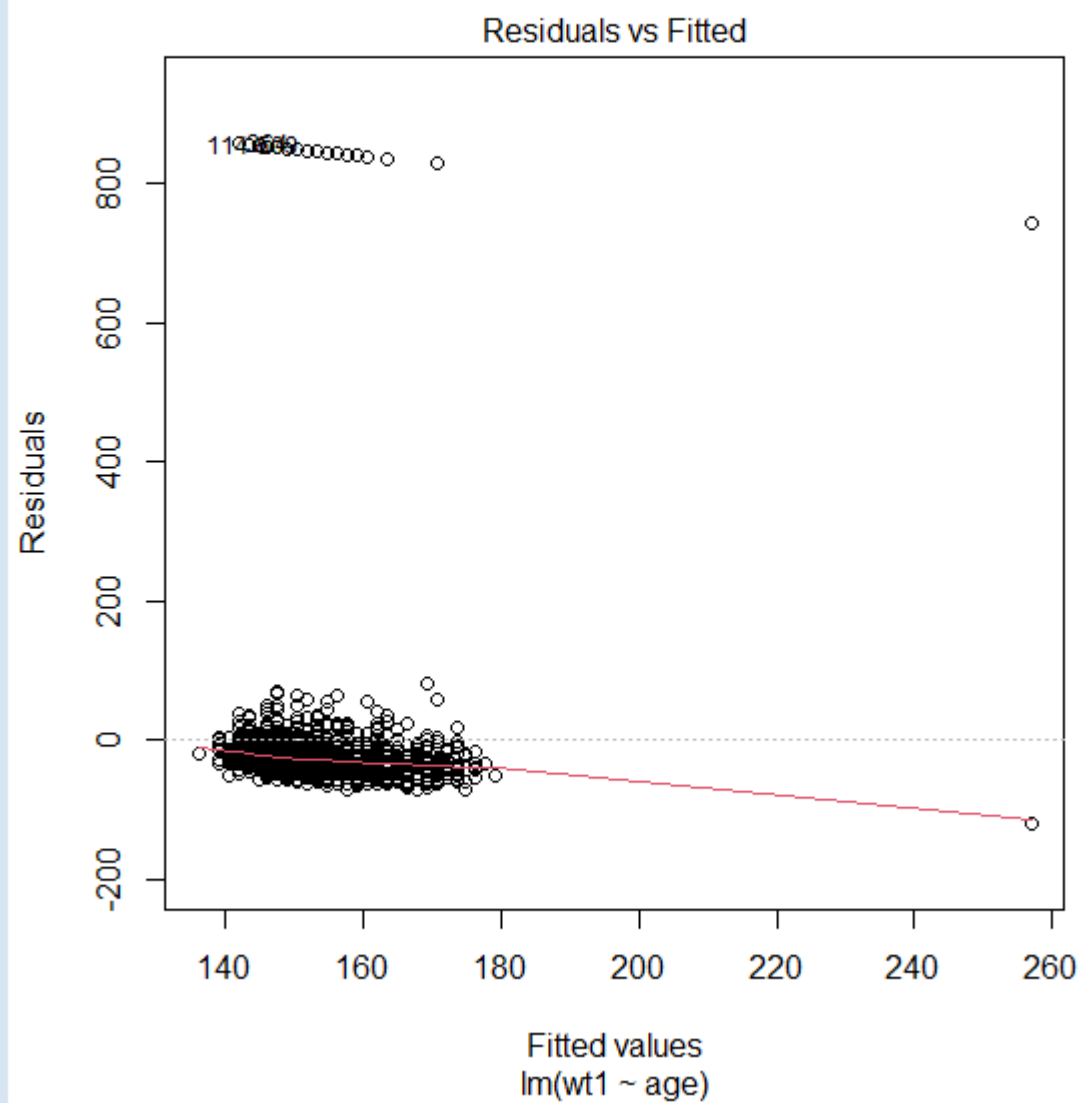
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

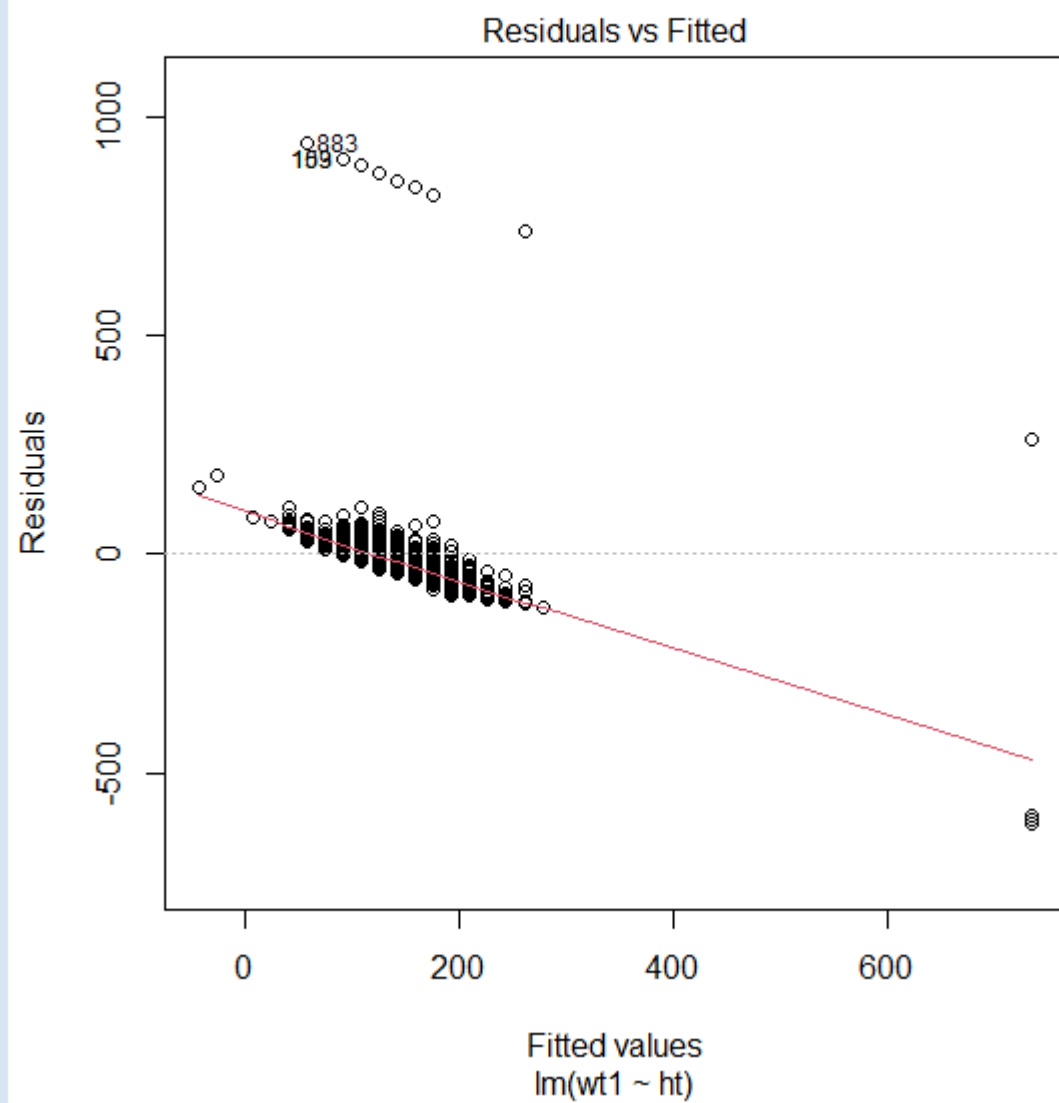
Residual standard error: 118.2 on 1234 degrees of freedom

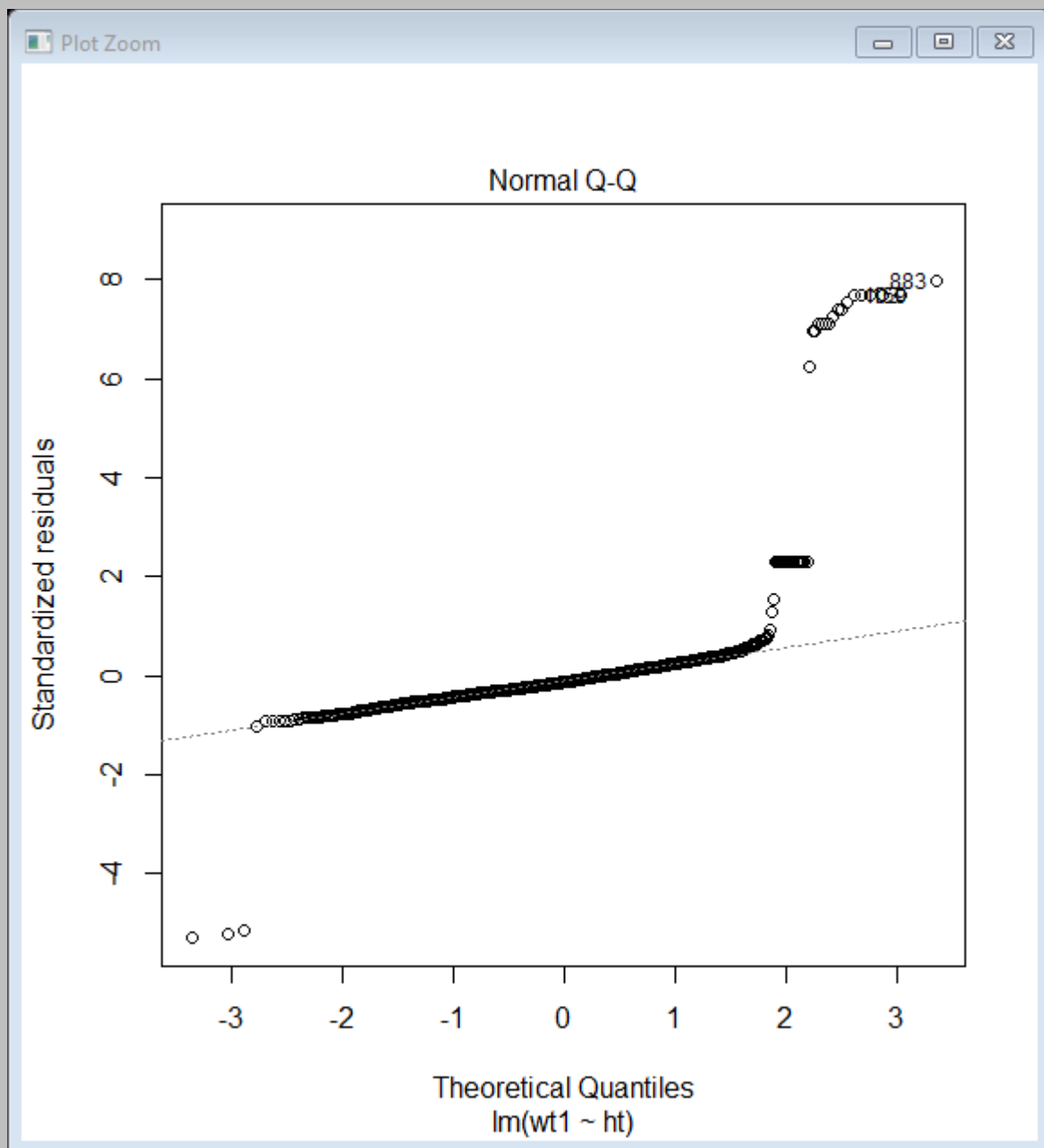
Multiple R-squared: 0.3612, Adjusted R-squared: 0.3607

F-statistic: 697.8 on 1 and 1234 DF, p-value: < 2.2e-16

```
> detach(babies)
```







Finish

