

資料探勘導論

10173241 張庭薰

目錄

- 目的與動機
- 方法與分析
- 結論
- 資料參考

目的與動機

想了解睡眠品質會因為哪種原因而受到影響及影響程度的高低，並運用資料及找出有哪些變因，及他們的程度高低。

方法與分析

資料尋找及匯入

ion 2023.1.R* x

data x

Regression 2023.1-4.R* x

iris x

Sleep_health_and_lifestyle_dataset x

»

←

→

🔍

Filter

🔍

	Person.ID	Gender	Occupation	Age	Sleep.Duration	Quality.of.Sleep	Physical.Activity.Level
1	1	Male	Software Engineer	27	6.1	6	42
2	2	Male	Doctor	28	6.2	6	60
3	3	Male	Doctor	28	6.2	6	60
4	4	Male	Sales Representative	28	5.9	4	30
5	5	Male	Sales Representative	28	5.9	4	30
6	6	Male	Software Engineer	28	5.9	4	30
7	7	Male	Teacher	29	6.3	6	40
8	8	Male	Doctor	29	7.8	7	75
9	9	Male	Doctor	29	7.8	7	75
10	10	Male	Doctor	29	7.8	7	75
11	11	Male	Doctor	29	6.1	6	30
12	12	Male	Doctor	29	7.8	7	75
13	13	Male	Doctor	29	6.1	6	30
14	14	Male	Doctor	29	6.0	6	30
15	15	Male	Doctor	29	6.0	6	30

匯入kaggle中尋找的csv資料

Sleep_health_and_lifestyle_dataset

並檢查內部資料

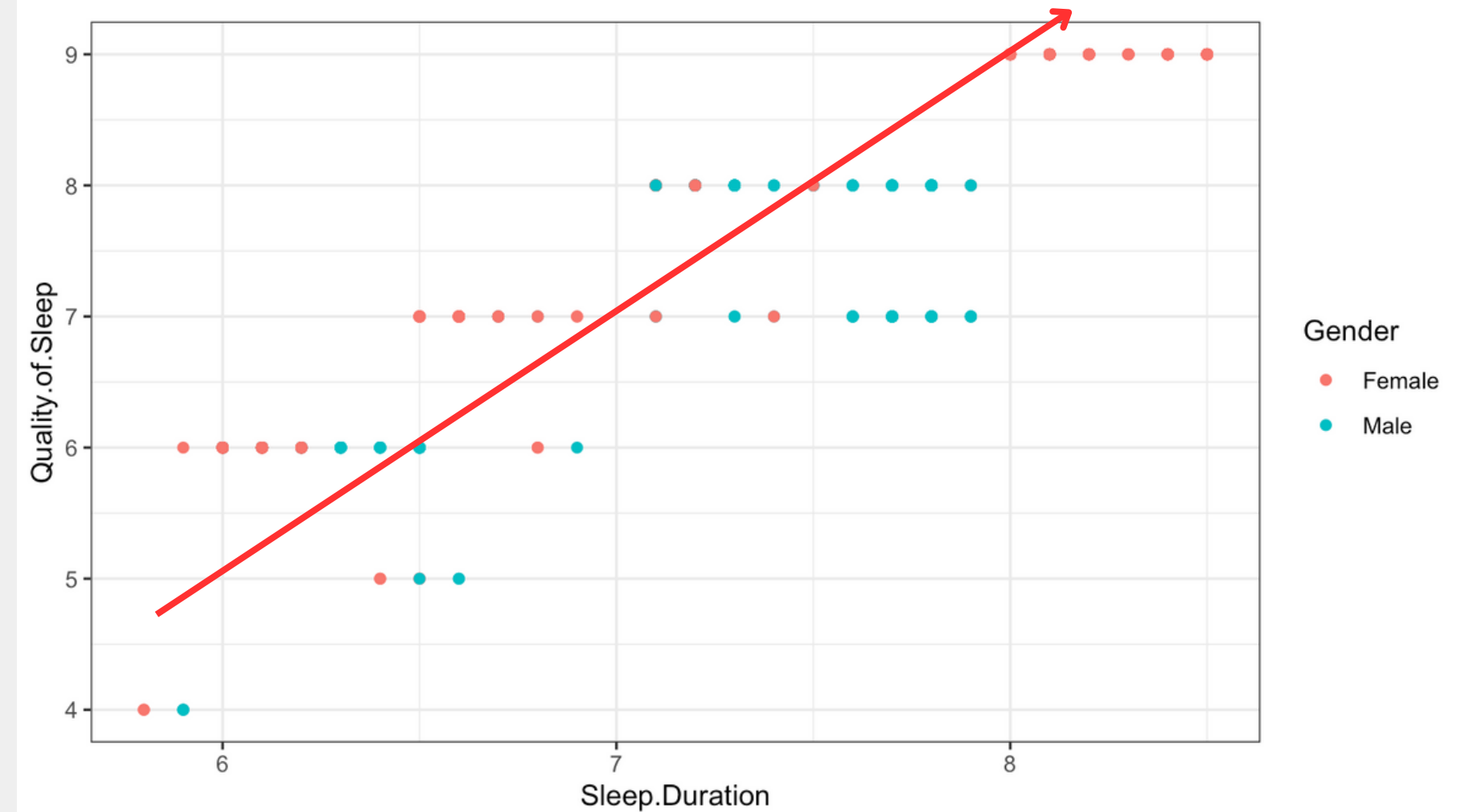
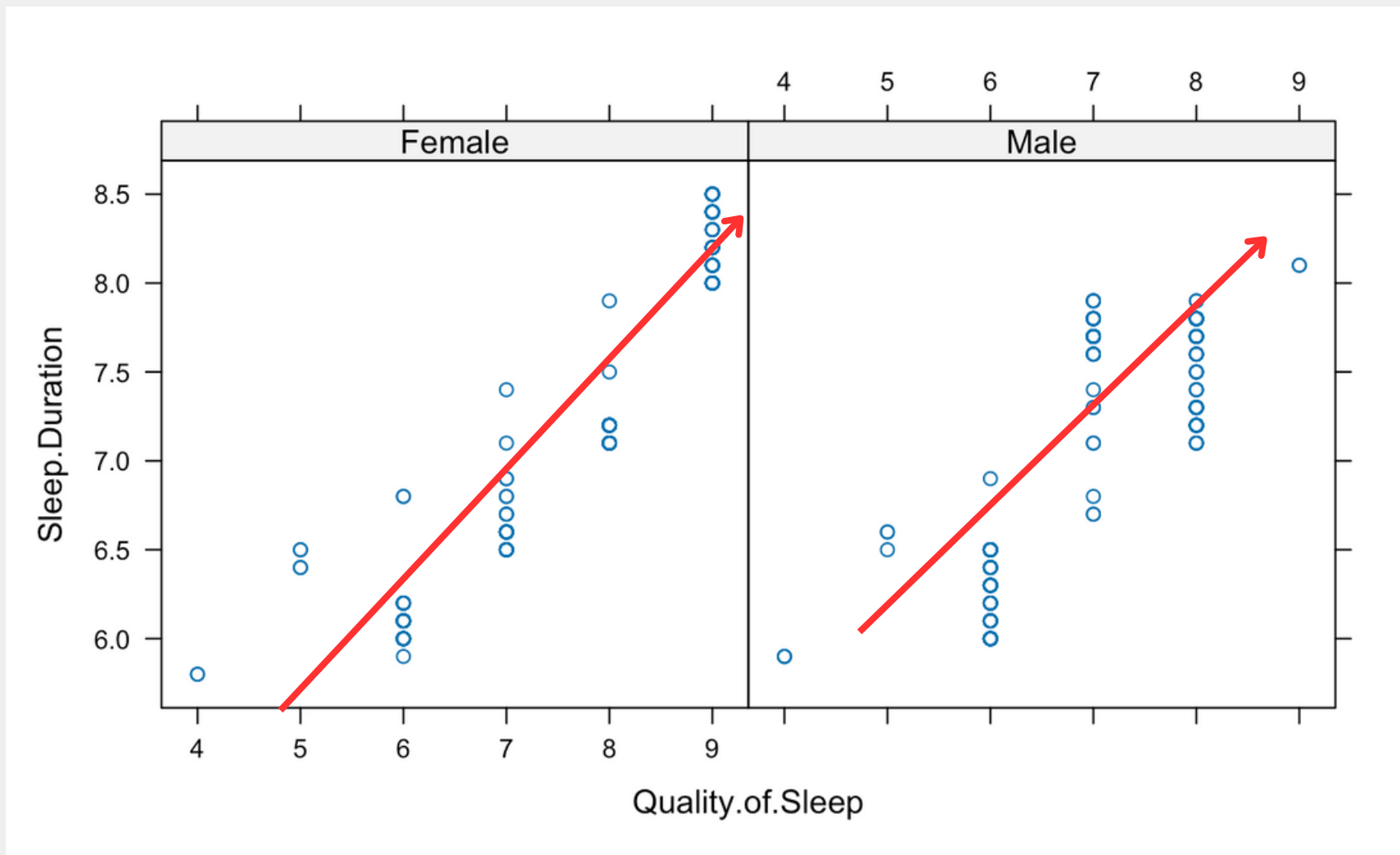
```
> summary(Sleep_health_and_lifestyle_dataset)
```

Person.ID	Gender	Occupation	Age
Min. : 1.00	Length:374	Length:374	Min. :27.00
1st Qu.: 94.25	Class :character	Class :character	1st Qu.:35.25
Median :187.50	Mode :character	Mode :character	Median :43.00
Mean :187.50		Mean :42.18	
3rd Qu.:280.75		3rd Qu.:50.00	
Max. :374.00		Max. :59.00	
Sleep.Duration	Quality.of.Sleep	Physical.Activity.Level	Stress.Level
Min. :5.800	Min. :4.000	Min. :30.00	Min. :3.000
1st Qu.:6.400	1st Qu.:6.000	1st Qu.:45.00	1st Qu.:4.000
Median :7.200	Median :7.000	Median :60.00	Median :5.000
Mean :7.132	Mean :7.313	Mean :59.17	Mean :5.385
3rd Qu.:7.800	3rd Qu.:8.000	3rd Qu.:75.00	3rd Qu.:7.000
Max. :8.500	Max. :9.000	Max. :90.00	Max. :8.000
BMI.Category	Blood.Pressure	Heart.Rate	Daily.Steps
Length:374	Length:374	Min. :65.00	Min. : 3000
Class :character	Class :character	1st Qu.:68.00	1st Qu.: 5600
Mode :character	Mode :character	Median :70.00	Median : 7000
		Mean :70.17	Mean : 6817
		3rd Qu.:72.00	3rd Qu.: 8000
		Max. :86.00	Max. :10000

方法與分析

繪圖分析

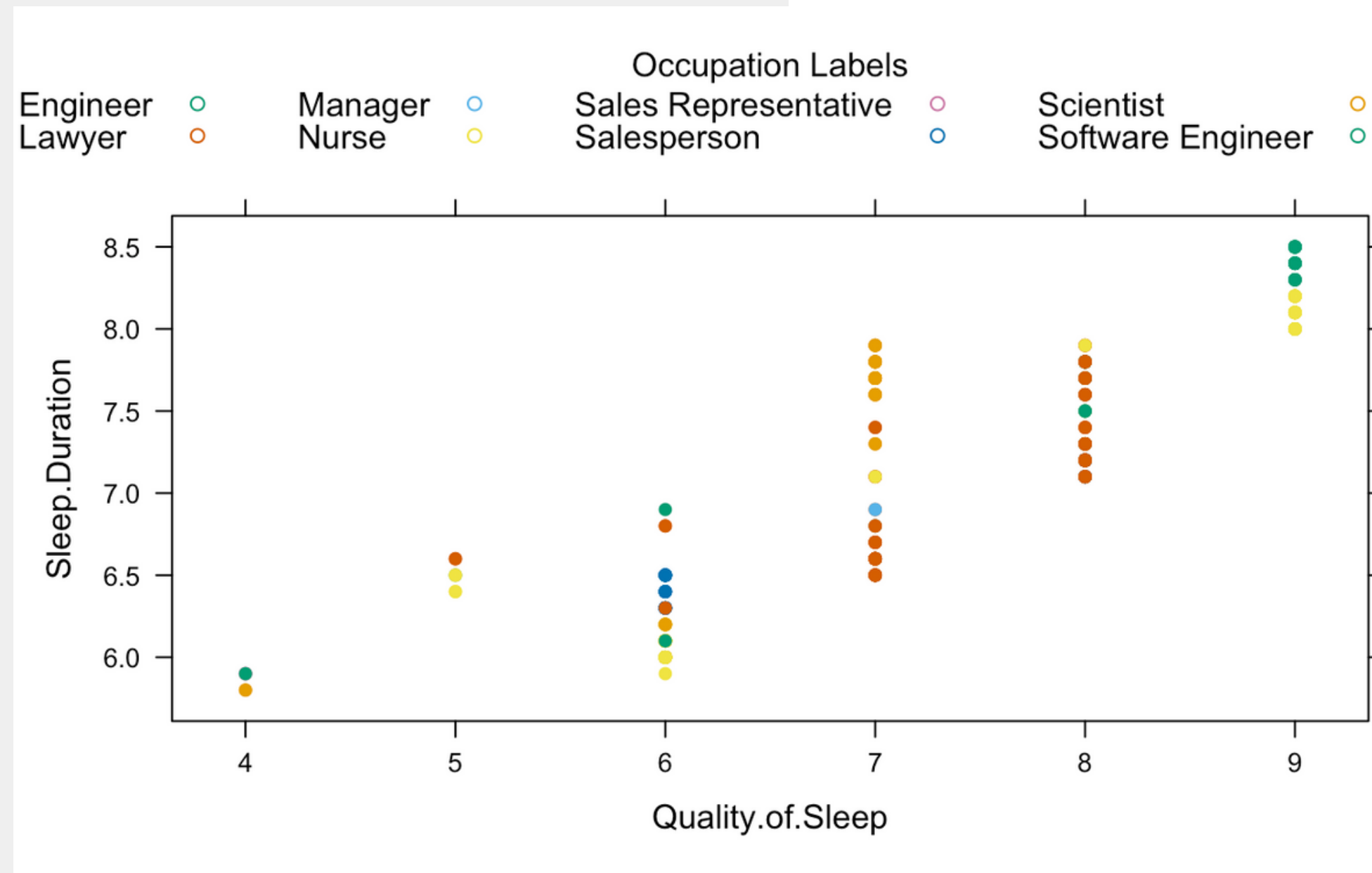
以兩圖做對比男女的睡眠品質都是因睡眠時常而有所改變
睡眠時常與品質成正相關，代表睡眠時間越長睡眠品質則越高。



方法與分析

繪圖分析

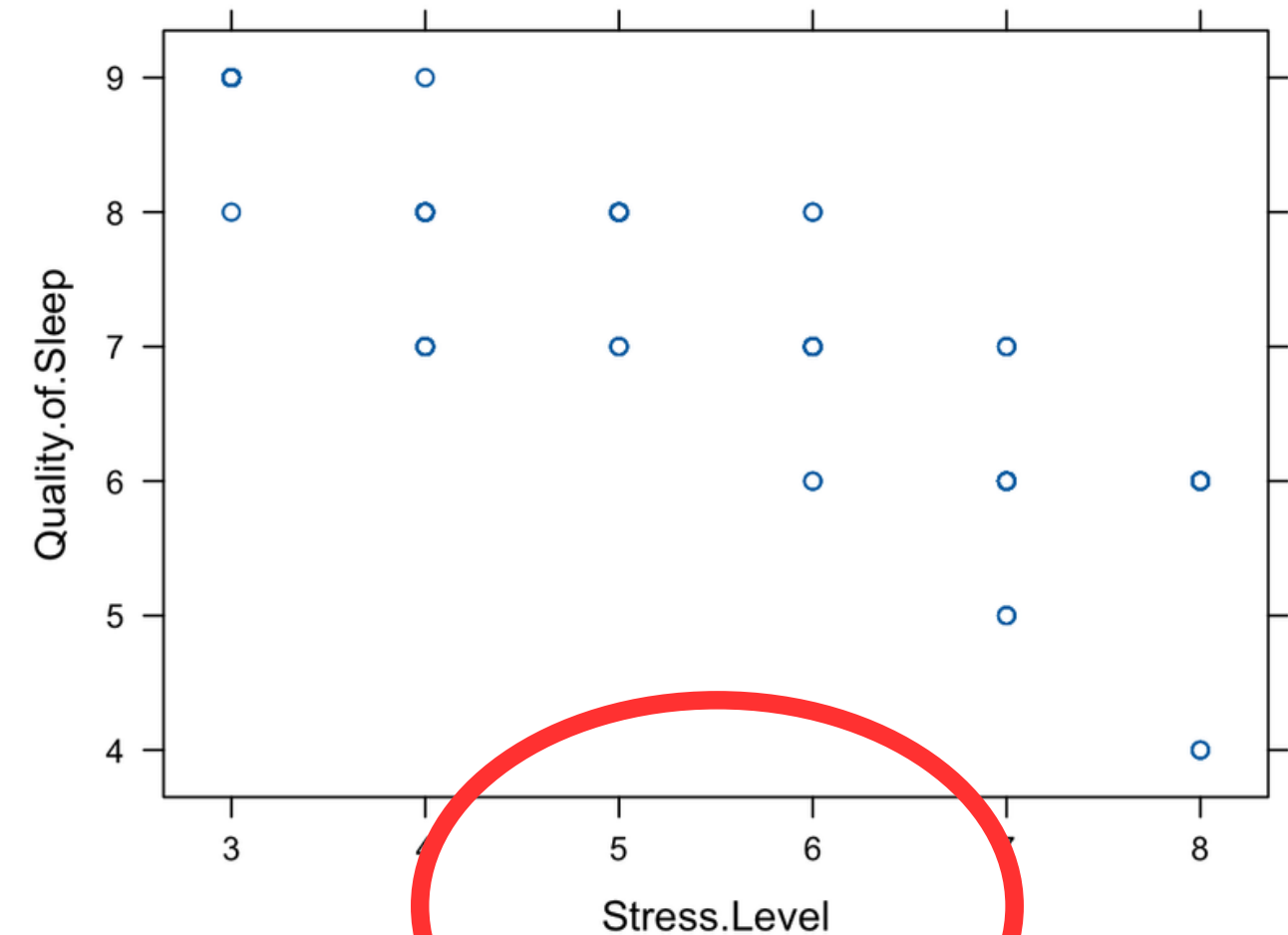
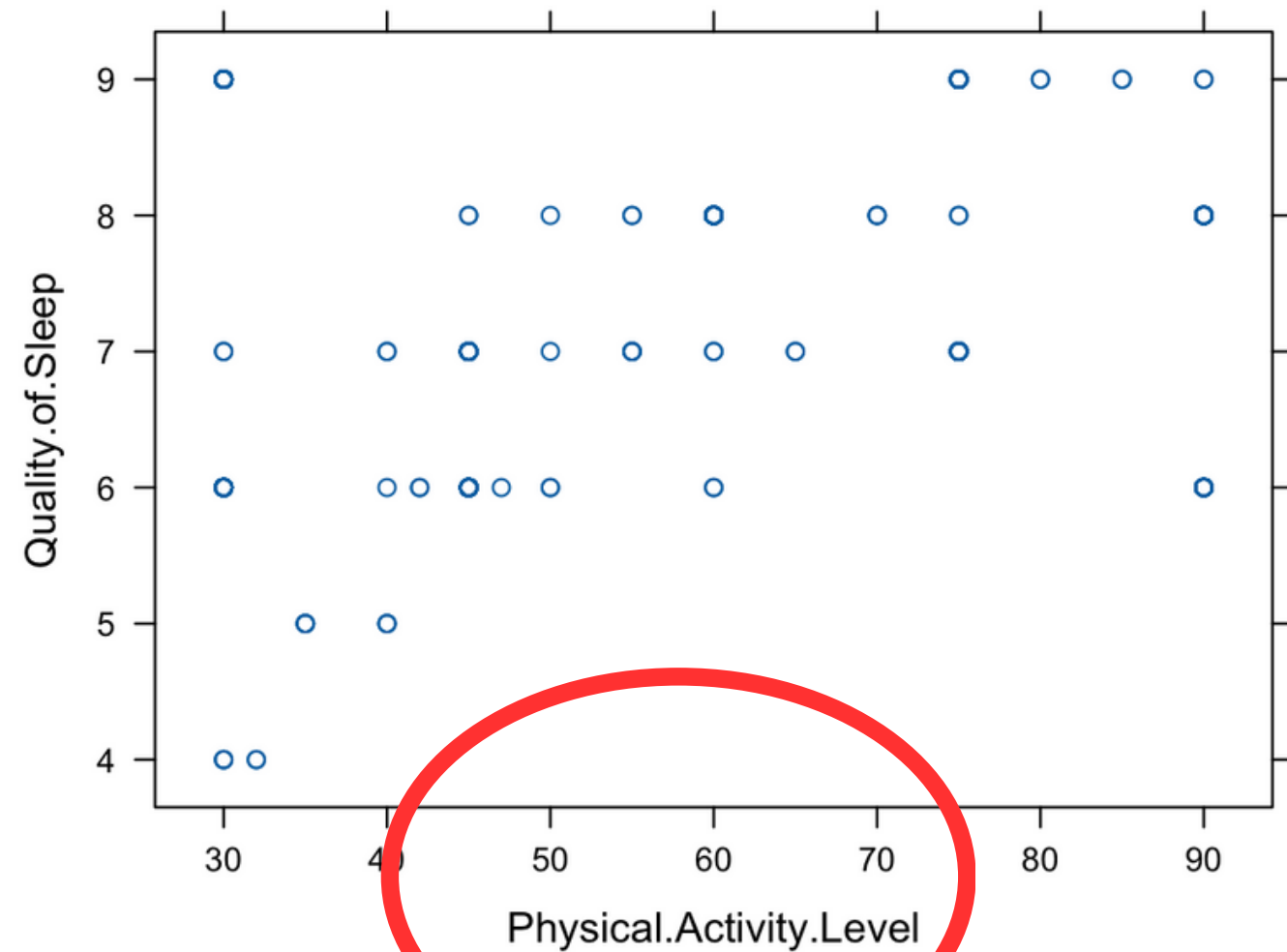
再對比所有職業的睡眠品質與睡眠期間呈正相關，
以此更能證明睡眠時間越長睡眠品質則越高。



方法與分析

繪圖分析

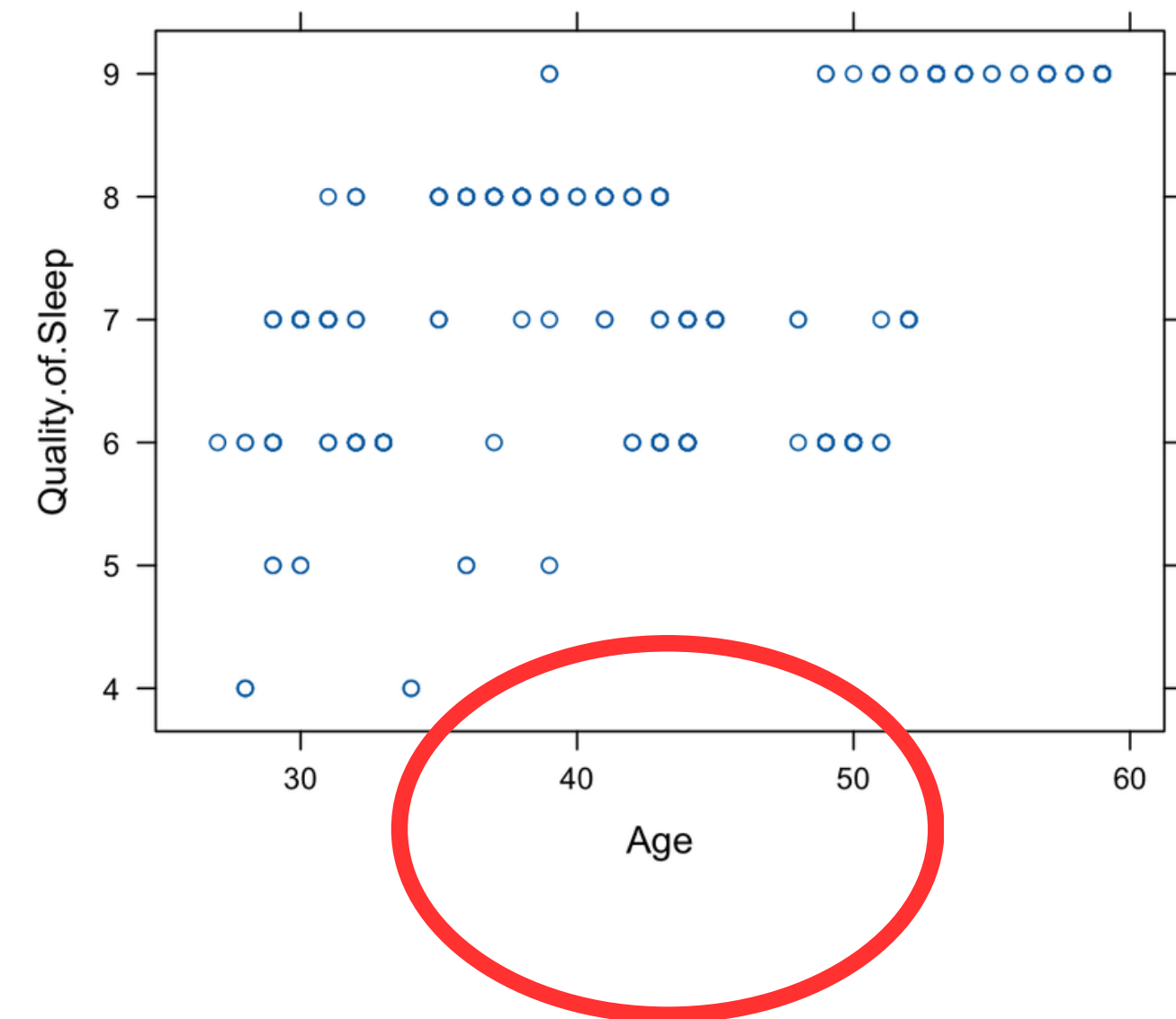
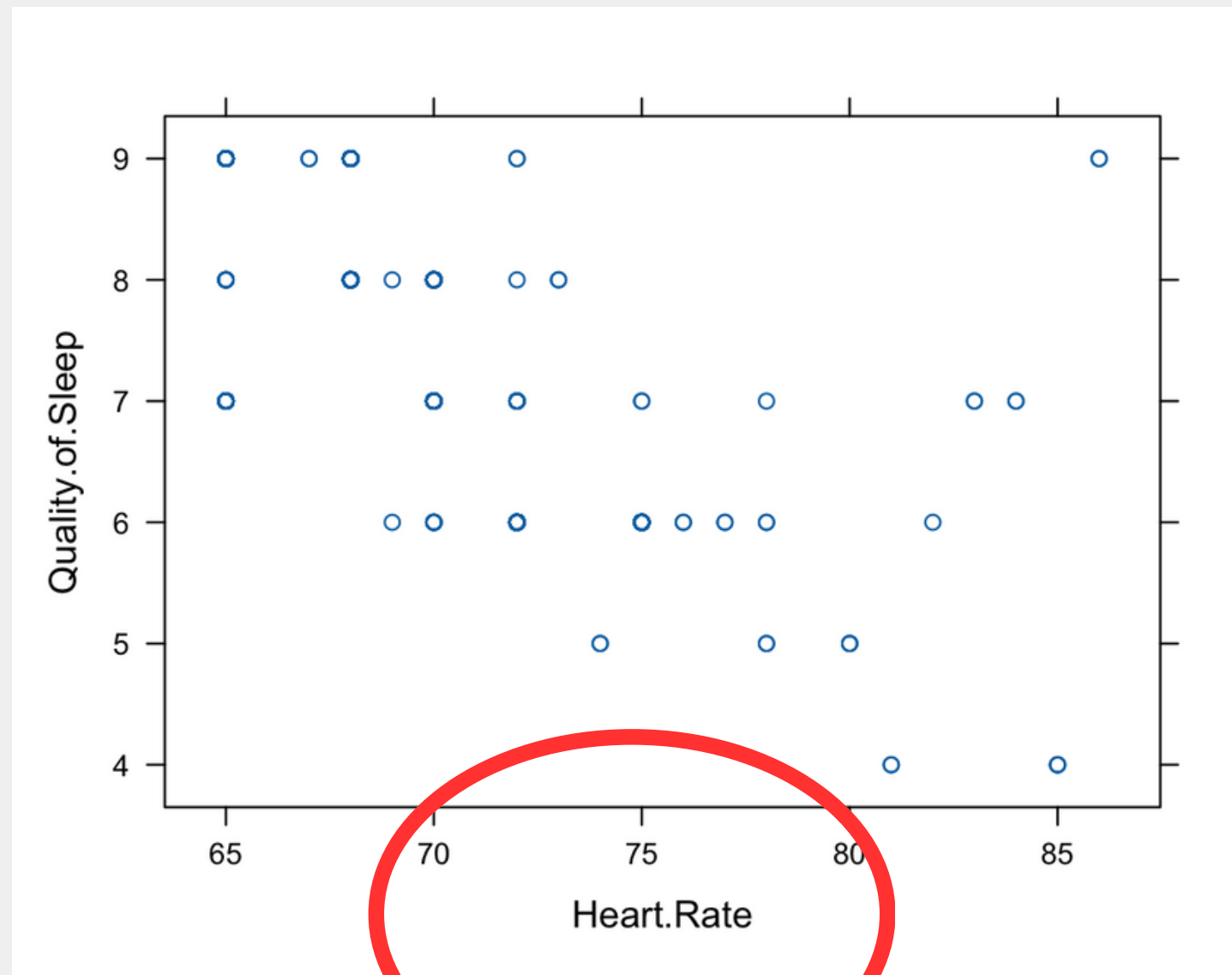
根據資料及做分析能了解身體活動程度及壓力等級都與睡眠品質有一定的關聯度。
由右圖壓力越大睡眠品質則越差，由左圖去除掉偏差值身體活動程度越大則睡眠品質越高。



方法與分析

繪圖分析

如右圖年齡越大則睡眠品質則越佳
如左圖心率越低則睡眠品質越佳



方法與分析

遺漏值處理

資料處理是否有遺漏值

```
> data <- data.frame(x=c(1,2,3,NA,5),
+                    y=c(4,5,3,NA,NA))
> data
  x y
1 1 4
2 2 5
3 3 3
4 NA NA
5 5 NA
> is.na(data)
      x      y
[1,] FALSE FALSE
[2,] FALSE FALSE
[3,] FALSE FALSE
[4,]  TRUE  TRUE
[5,] FALSE  TRUE
> table(is.na(data)) #有多少個遺漏值

FALSE TRUE
    7    3
```

```
> # 「移除有遺漏值得資料」 & 「填補遺漏值」
> # 移除有遺漏值得資料
> data[complete.cases(data), ]
  x y
1 1 4
2 2 5
3 3 3
> na.omit(data)
  x y
1 1 4
2 2 5
3 3 3
>
> # 填補遺漏值 (用平均數填值)
> data[is.na(data[, "y"]), "y"] <- mean(data[, "y"], na.rm=T)
> data
  x y
1 1 4
2 2 5
3 3 3
4 NA 4
5 5 4
>
> table(is.na(Sleep_health_and_lifestyle_dataset))

FALSE
4488
```

方法與分析

迴歸分析

根據線性迴歸模型，睡眠時長每增加一單位，睡眠品質平均提高約0.61個單位。同樣，壓力水平每增加一單位，睡眠品質平均下降約0.38個單位。此外，每增加一單位的身體活動水平，睡眠品質平均提高約0.005個單位。模型的R-squared值為0.8843，表示約88.43%的睡眠品質變異可以由模型中的變數解釋。這些結果均具有統計顯著性（ $p\text{-value} < 0.05$ ）。

```
Call:
lm(formula = Quality.of.Sleep ~ Sleep.Duration + Stress.Level +
    Physical.Activity.Level, data = Sleep_health_and_lifestyle_dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-1.41501 -0.23836  0.09995  0.20096  0.92804

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.750274   0.425059  11.176 < 2e-16 ***
Sleep.Duration      0.606984   0.047935  12.663 < 2e-16 ***
Stress.Level     -0.383480   0.021015 -18.248 < 2e-16 ***
Physical.Activity.Level 0.005045   0.001072   4.708 3.55e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4089 on 370 degrees of freedom
Multiple R-squared:  0.8843, Adjusted R-squared:  0.8833
F-statistic: 942.3 on 3 and 370 DF, p-value: < 2.2e-16
```

結論

由以上分析可得出睡眠品質會因許多原因種而受到影響：
睡眠期間（討論的重點）、年齡、壓力、身體活動程度，
並運用以上方法證明其有相關性。

參考資料

Sleep Health and Lifestyle Dataset

THANK YOU