

데이터 분석 프로그래밍 중간고사

coop711

2018-4-18

학번 _____

이름 _____

BMI

수업시간에 공부한 rn96 에 대하여 다음 물음에 답하시오.

1. 다음 중 rn96 을 읽어들이고 내용을 출력하는 코드로 적절치 않은 것은?

가. rn96 <- read.table("../data/rn96.txt", header = TRUE);rn96

나. (rn96 <- read.table("../data/rn96.txt", header = TRUE))

다. assign(rn96, read.table("../data/rn96.txt", header = TRUE));rn96

라. assign("rn96", read.table("../data/rn96.txt", header = TRUE));rn96

2. str(rn96) 의 결과가 다음 글상자의 내용과 같다고 했을 때, 보기 중에서 그 결과가 다른 것 하나를 고르시오.

가. rn96\$height

나. rn96[, 1]

다. rn96[, "height"]

라. rn96["height"]

```
'data.frame':  41 obs. of  2 variables:
 $ height: int  161 155 158 170 160 156 162 158 158 167 ...
 $ weight: int  50 49 42 65 60 52 58 46 45 51 ...
```

3. 위의 rn96 에 대하여 각 변수의 산술평균을 한번에 구하고자 할 때 결과가 다른 구조로 나오는 것을 하나 고르시오.

가. sapply(rn96, mean)

나. apply(rn96, 2, mean)

다. mapply(mean, rn96)

라. lapply(rn96, mean)

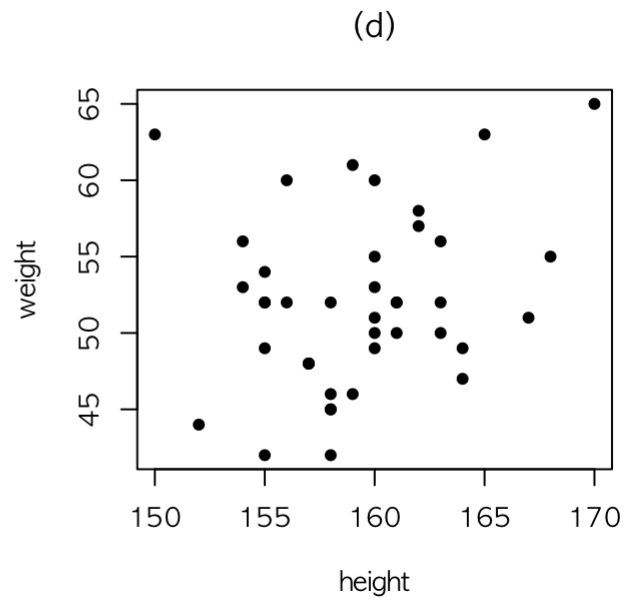
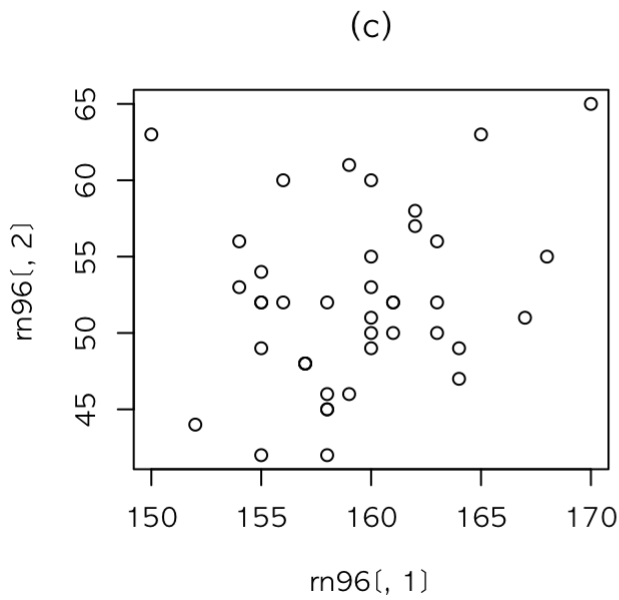
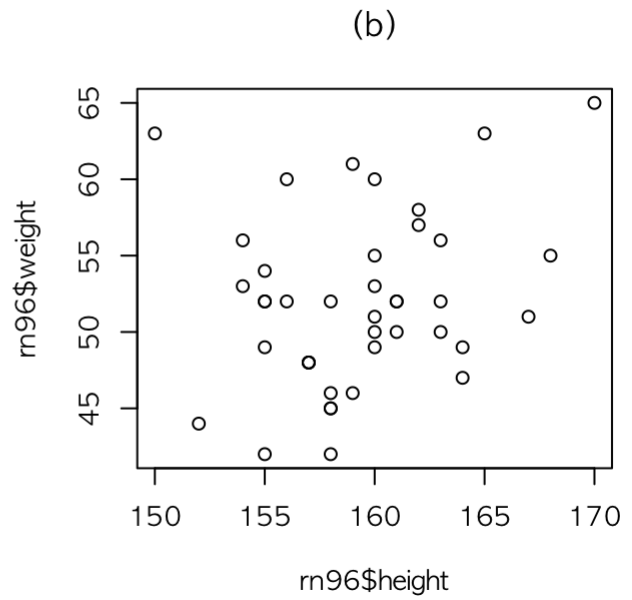
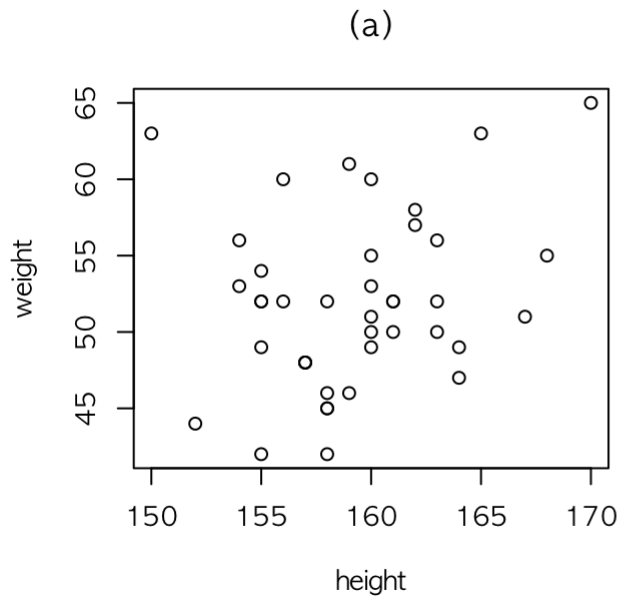
4. 여러가지 방법으로 산점도를 그려보았다. 어느 코드로 어느 산점도를 그린 것인지 짚 지으시오.

가. plot(x = rn96[, 1], y = rn96[, 2])

나. plot(rn96\$height, rn96\$weight)

다. plot(weight ~ height, data = rn96)

라. plot(rn96, pch = 16)



5. 위와 같이 4개의 그림을 2×2 행렬 모양으로 배치하려면 어떤 설정이 우선 필요한가?

가. `par(mfrow = c(2, 2))`

나. `options(mfrow = c(2, 2))`

다. `par(nrow = 2, ncol = 2)`

라. `options(nrow = 2, ncol = 2)`

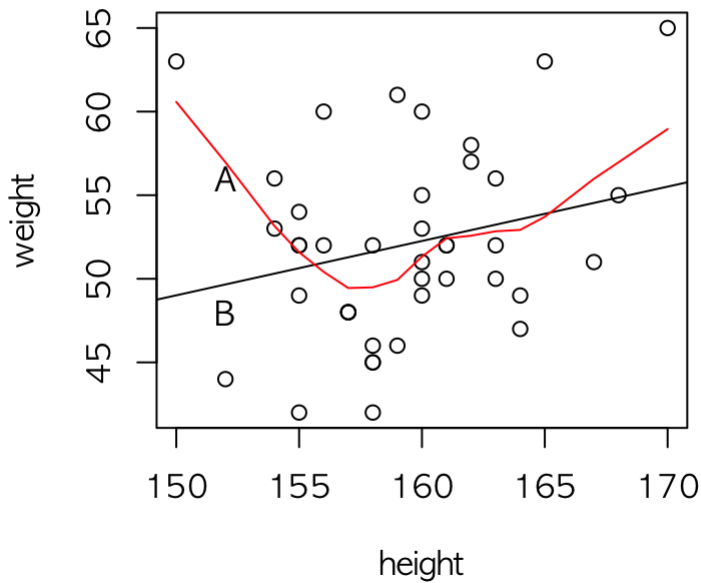
6. 다음과 같이 산점도에 두 종류의 회귀선을 그리는 데 사용된 함수를 제대로 짚지은 것은?

가. A: `lm`, B: `lowess`

나. A: `lowess`, B: `lm`

다. A: `lines`, B: `points`

라. A: `points`, B: `lines`



7. BMI 를 토대로 한 비만도 판정은 “18.5 미만은 underweight, 18.5 ~ 24.9 는 Normal, 25 ~ 29.9 는 Overweight, 30 이상은 Obese” 이다. 이 때, rn96 에 판정결과를 덧붙이기 위하여 작성한 다음 코드에서 괄호 자리에 공통으로 들어가는 R 함수는 무엇인가?

가. ifelse

나. if

다. for

라. while

```
rn96$BMI <- BMI
(rn96$obesity <- ( (1) )(BMI < 18.5, "Underweight",
                  ( 2) )(BMI >= 18.5 & BMI < 24.9, "Normal",
                  ( 3) )(BMI >= 25 & BMI < 29.9, "Overweight", "Obese"))))
```

8. 당초 height 와 finger 로 구성되어 있던 rn96 은 `위의 코드를 실행하고 나면 몇 개의 열로 구성되는가?

가. 2

나. 3

다. 4

라. 5

9. BMI 값이 24인 사람의 경우 번호로 표시한 (1), (2), (3) 중 어느 단계가 수행되는가?

가. (1)

나. (2)

다. (3)

라. 어느 것도 아니다

Quetelet's Scottish Soldiers

10. 다음과 같이 그림 파일로 저장되어 있는 자료를 불러들이는 코드로 적절한 것은? 단, 그림의 폭을 기준으로 적절히 조절하고자 한다.

가. `![Quetelet's frequency table](../pics/quetelet_soldiers.png)`

나. ``

다. `![Quetelet's frequency table](../pics/quetelet_soldiers.png, width = 160)`

라. ``

MEASURE de la taille	NOMBRE d'hommes	NOMBRE normalisés	PROBABILITÉ d'être l'un d'eux	RANG des tailles	RANG d'après la taille	PROBABILITÉ d'être l'un d'eux	NOMBRE d'hommes normalisés
33	5	5	0,0000			0,0000	7
34	18	21	0,0095	32	50	0,0095	29
35	81	141	0,0464	42,5	42,5	0,0464	119
36	185	329	0,0895	52,5	54,5	0,0894	235
37	420	732	0,4501	95,0	96,5	0,4501	732
38	749	1305	0,5709	18,0	18,5	0,5709	1335
39	1073	1907	0,5814	10,5	10,5	0,5814	1838
			0,5997	9,5	9,5	0,5998	
40	1079	1882	0,1985	5,5	5,5	0,1930	1987
41	954	1028	0,2915	13	13,5	0,2854	1075
42	608	1148	0,4061	21	21,5	0,4130	1095
43	370	665	0,4790	30	30,5	0,4809	560
44	92	160	0,4866	35	37,5	0,4911	291
45	50	87	0,4955	41	45,5	0,4989	69
46	21	38	0,4991	49,5	55,5	0,4995	16
47	4	7	0,4998	50	61,8	0,4999	5
48	1	2	0,5000			0,5000	1
	5728	1,0000					1,0000

11. 아래와 같은 구조를 가진 `chest.table` 에서 "Freq" 를 추출하는 방법 중 구조가 나머지 셋과 다른 것은?

가. `chest.table$Freq`

나. `chest.table[, 2]`

다. `chest.table[, "Freq"]`

라. `chest.table["Freq"]`

```
## 'data.frame':   16 obs. of  2 variables:
##  $ Chest: int   33 34 35 36 37 38 39 40 41 42 ...
##  $ Freq : num   3 18 81 185 420 ...
```

12. 다음 `barplot` 에서 막대 사이의 간격을 없애려면 어떤 조건을 넣어야 하는가?

가. `names.arg = 33:48`

나. `space = 0`

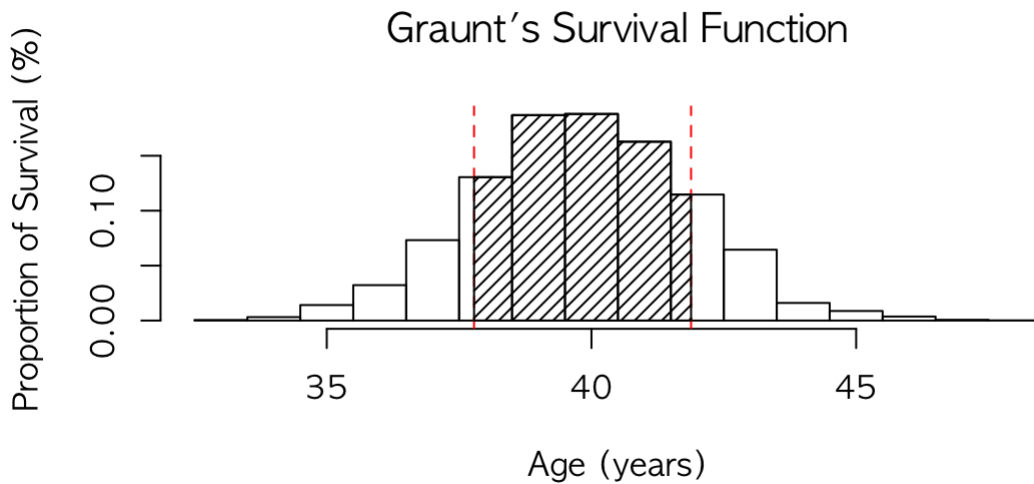
다. `offset = 0`

라. `beside = TRUE`



13. 아래 그림에서 빗금친 부분을 나타내기 위해서 사용한 R 함수는 무엇인가?

- 가. `plot`
- 나. `lines`
- 다. `abline`
- 라. `polygon`



14. 위의 그림에서 수직으로 세운 두 점선을 긋기 위해서 사용한 R 함수는 무엇인가?

- 가. `plot`
- 나. `lines`
- 다. `abline`
- 라. `polygon`

15. 위의 그림에서 빗금의 방향을 바꾸려면 다음 중 어느 조건을 설정하여야 하는가?

- 가. `angle =`
- 나. `density =`
- 다. `col =`
- 라. `border =`

16. 위의 `chest.table` 데이터 프레임을 33인치 세번, 34인치가 18번 등으로 반복해서 나오는 한 줄의 벡터로 바꾸려면 어떤 방법이 적절한가?

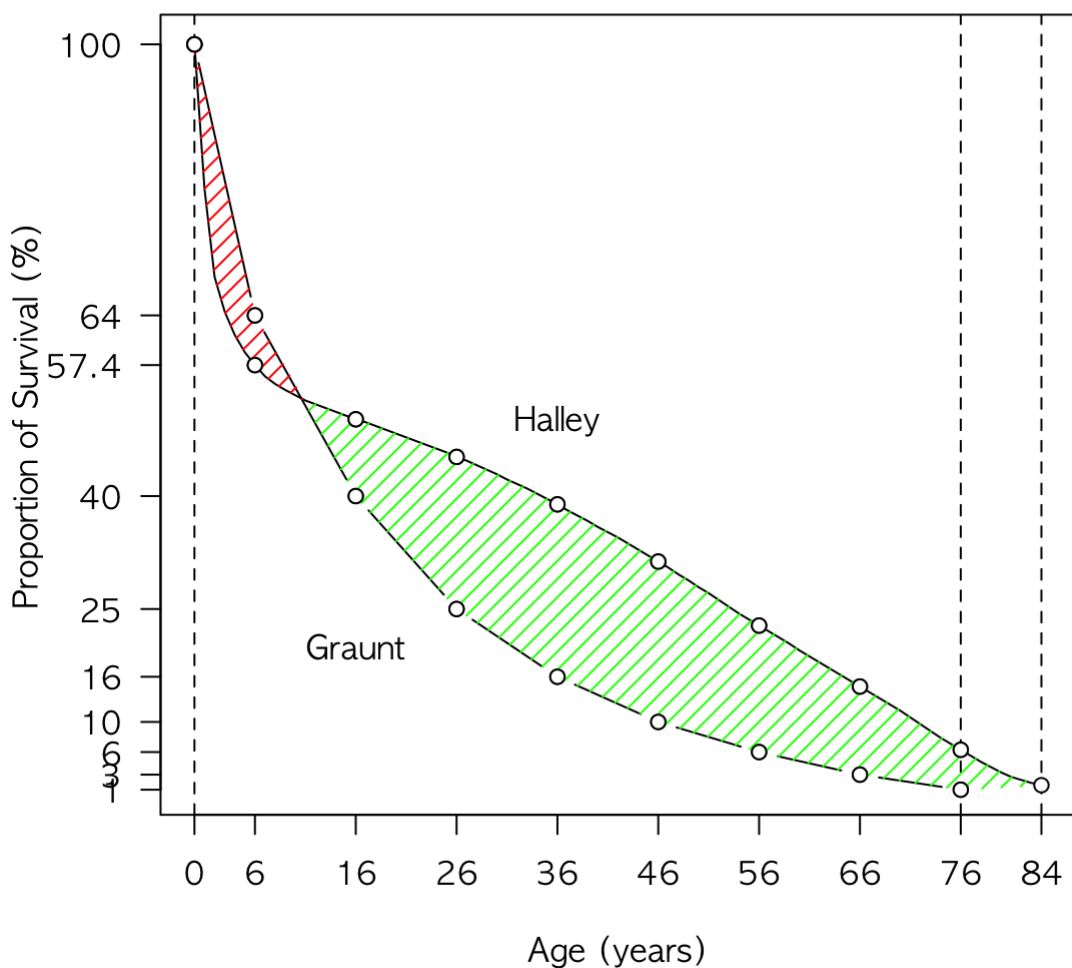
- 가. `rep(chest.table$Chest, chest.table$Freq)`
- 나. `rep(chest.table$Freq, chest.table$Chest)`
- 다. `rep(chest.table$Chest, each = chest.table$Freq)`
- 라. `rep(chest.table$Freq, each = chest.table$Chest)`

Lifetable

다음 소스코드와 출력결과물을 보고 물음에 답하시오.

```
plot(halley, ann = FALSE, xaxt = "n", yaxt = "n", type = "l")
abline(v = c(0, 76, 84), lty = 2)
points(halley.graunt, pch = 21, col = "black", bg = "white")
lines(graunt, type = "b", pch = 21, col = "black", bg = "white")
axis(side = 1, at = c(graunt$x, 84), labels = c(graunt$x, 84))
axis(side = 2, at = graunt$xPo.g, labels = graunt$xPo.g, las = 1)
axis(side = 2, at = xPo.halley.age.6, labels = xPo.halley.age.6, las = 1)
text(x = c(16, 36), y = c(20, 50), label = c("Graunt", "Halley"))
title(main = main.title.2, xlab = x.lab, ylab = y.lab)
polygon(poly.upper, angle = 45, density = 15, col = "red", border = NA)
polygon(poly.lower, angle = 45, density = 15, col = "green", border = NA)
points(graunt, pch = 21, col = "black", bg = "white")
points(halley.graunt, pch = 21, col = "black", bg = "white")
points(x = 84, y = halley$xPo[85], pch = 21, col = "black", bg = "white")
```

Survival Function of Graunt and Halley



17. 이 중에서 도표의 제목을 사용자가 입력하기 위해서 집어넣은 조건은 무엇인가?

가. `ann = FALSE`

나. `xaxt = "n"`

다. `yaxt = "n"`

라. `type = "l"`

18. 관찰 연령의 시작과 끝을 뚜렷이 나타내기 위해서 사용한 코드는 무엇인가?

가. `abline(v = c(0, 76, 84), lty = 2)`

나. `axis(side = 1, at = c(graunt$x, 84), labels = c(graunt$x, 84))`

다. `axis(side = 2, at = graunt$xPo.g, labels = graunt$xPo.g, las = 1)`

라. `axis(side = 2, at = xPo.halley.age.6, labels = xPo.halley.age.6, las = 1)`

19. y축 눈금의 숫자를 축과 직각이 되도록 하는 코드는 무엇인가?

가. `side = 2`

나. `at = graunt$xPo.g`

다. `labels = graunt$xPo.g`

라. `las = 1`

20. 오른쪽 y축에 눈금과 값을 표시하려면 어떤 조건이 필요한가?

가. `side = 3`

나. `side = 4`

다. `las = 1`

라. `las = 2`

21. Halley 생존표에서만 관찰되는 6세 연령의 생존률을 y축에 표시하기 위하여 작성된 코드는 무엇인가?

가. `abline(v = c(0, 76, 84), lty = 2)`

나. `axis(side = 1, at = c(graunt$x, 84), labels = c(graunt$x, 84))`

다. `axis(side = 2, at = graunt$xPo.g, labels = graunt$xPo.g, las = 1)`

라. `axis(side = 2, at = xPo.halley.age.6, labels = xPo.halley.age.6, las = 1)`

22. 빗금친 부분을 표시하는 코드에서 굳이 넣지 않아도 되는 것은 무엇인가?

가. `angle = 45`

나. `density = 15`

다. `col = "green"`

라. `border = NA`

23. 이 코드 중에서 점의 윤곽을 뚜렷이 하기 위하여 작성된 부분은 어디인가?

가. `abline(v = c(0, 76, 84), lty = 2)`

나. `axis(side = 2, at = xPo.halley.age.6, labels = xPo.halley.age.6, las = 1)`

다. `text(x = c(16, 36), y = c(20, 50), label = c("Graunt", "Halley"))`

라. `points(graunt, pch = 21, col = "black", bg = "white")`

24. Halley의 기대수명과 Graunt의 기대수명의 차이는 어떻게 나타나는가?

- 가. 윗쪽 빗금친 면적과 아랫쪽 빗금친 면적의 합
- 나. 윗쪽 빗금친 면적에서 아랫쪽 빗금친 면적을 뺀 값
- 다. 아랫쪽 빗금친 면적에서 윗쪽 빗금친 면적을 뺀 값
- 라. 두 곡선의 아랫 부분 면적

25. 아래의 reshaping 과정에서 factor를 설정하는 부분은 어디인가?

- 가. `list(graunt.2, halley.2, us93.2)`
- 나. `id.vars = "x"`
- 다. `value.name = "xPo"`
- 라. `variable.name = "Who"`

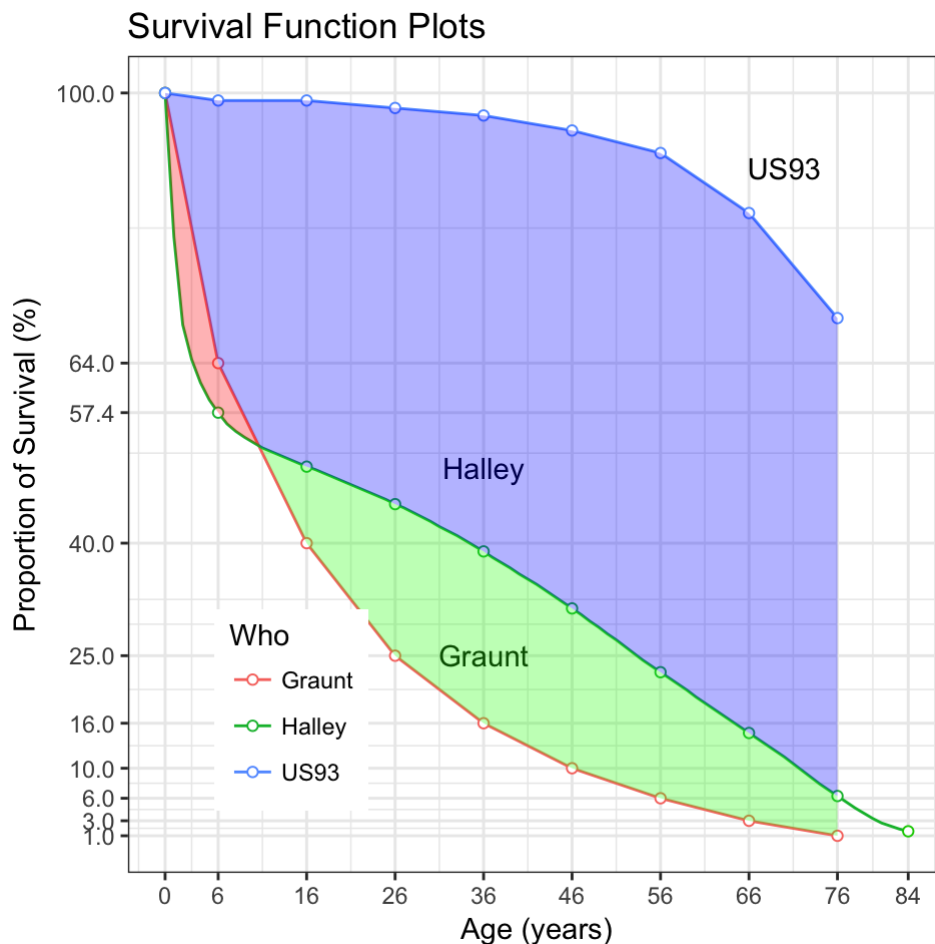
26. 이 데이터 변환 작업에 필요한 패키지는 무엇인가?

- 가. `ggplot2`
- 나. `ggplot`
- 다. `reshape2`
- 라. `reshape`

```
ghu.melt <- melt(list(graunt.2, halley.2, us93.2), id.vars = "x", value.name = "xPo",  
  variable.name = "Who")
```


다음 소스코드와 출력결과물을 보고 물음에 답하시오.

```
(ghup <- ggplot() +  
  geom_line(data = ghu.melt, aes(x = x, y = xPo, colour = Who)) +  
  geom_point(data = ghu.melt.g, aes(x = x, y = xPo, colour = Who), shape = 21, fill =  
"white") +  
  theme_bw() +  
  xlab(x.lab) +  
  ylab(y.lab) +  
  ggtitle(main.title.3) +  
  theme(legend.position = c(0.2, 0.2)) +  
  annotate("text", x = c(36, 36, 70), y = c(25, 50, 90), label = c("Graunt", "Halley"  
, "US93")) +  
  scale_x_continuous(breaks = c(graunt$x, 84)) +  
  scale_y_continuous(breaks = c(graunt$xPo.g, xPo.halley.age.6)) +  
  geom_polygon(data = poly.upper, aes(x = x, y = y), alpha = 0.3, fill = "red") +  
  geom_polygon(data = poly.lower.76, aes(x = x, y = y), alpha = 0.3, fill = "green")  
+  
  geom_polygon(data = poly.us.76, aes(x = x, y = y), alpha = 0.3, fill = "blue") +  
  geom_point(data = data.frame(x = 84, y = halley$xPo[85]), aes(x = x, y = y), colour =  
"red", shape = 21, fill = "white"))
```



27. 뼈대가 되는 생존곡선을 그리는 과정에 각 집단을 구분하는 코드는 무엇인가?

가. `x = x, y = xPo`

나. `colour = Who`

다. `shape = 21`

라. `fill = "white"`

28. 도표의 배경을 흑백으로 바꿔주는 코드는 ?

가. `theme_bw()`

나. `xlab(x.lab)`

다. `ylab(y.lab)`

라. `ggtitle(main.title.3)`

29. 범례를 도표 안쪽으로 위치시키는 코드는 무엇인가?

가. `theme(legend.position = c(0.2, 0.2))`

나. `theme_bw()`

다. `scale_x_continuous(breaks = c(graunt$x, 84))`

라. `scale_y_continuous(breaks = c(graunt$xPo.g, xPo.halley.age.6))`

30. R의 Base Plot에서 `axis(side = 2, ...)` 와 같은 기능을 수행하는 코드는 무엇인가?

가. `theme(legend.position = c(0.2, 0.2))`

나. `theme_bw()`

다. `scale_x_continuous(breaks = c(graunt$x, 84))`

라. `scale_y_continuous(breaks = c(graunt$xPo.g, xPo.halley.age.6))`

31. 색깔 투명도를 조절하는 코드는 무엇인가?

가. `alpha = 0.3`

나. `colour = Who`

다. `shape = 21`

라. `fill = "white"`

32. 도표 제목의 위치를 가운데로 옮기려면 `theme()` 안에 어떤 코드가 추가로 필요한가?

가. `, plot.title = element_rect(hjust = 0.5)`

나. `, plot.title = element_text(hjust = 0.5)`

다. `, plot.title = element_rect(vjust = 0.5)`

라. `, plot.title = element_text(vjust = 0.5)`

33. `xlab(x.lab) + ylab(y.lab) + ggtitle(main.title.3)` 를 하나의 함수로 표현하면?

가. `labs(x = x.lab, y = y.lab, main = main.title.3)`

나. `labs(x = x.lab, y = y.lab, title = main.title.3)`

다. `labs(xlab = x.lab, ylab = y.lab, main = main.title.3)`

라. `labs(xlab = x.lab, ylab = y.lab, title = main.title.3)`