

중간고사

통계적 방법론

2016-4-22

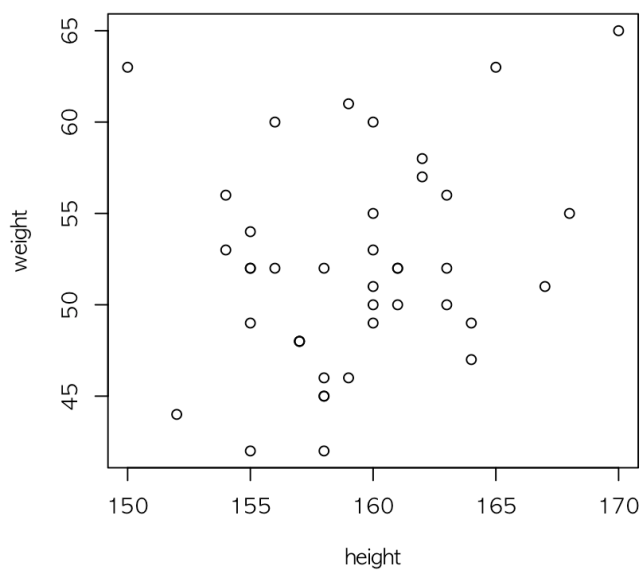
학번 _____

이름 _____

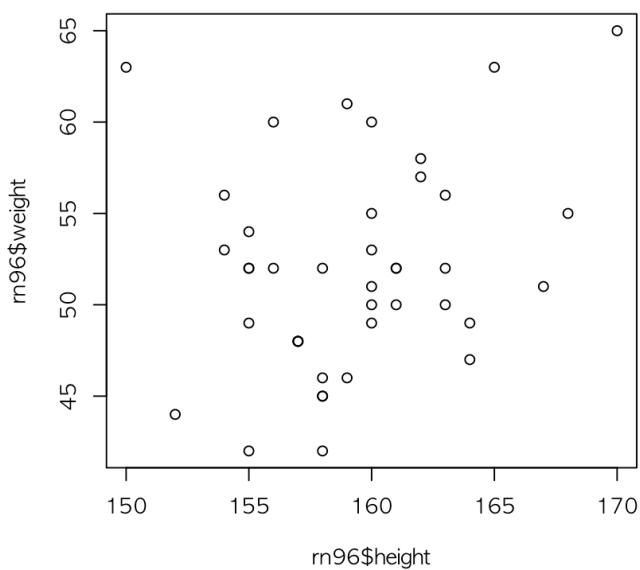
BMI

1. 다음 중 `plot(rn96$height, rn96$weight)` 의 결과로 적합한 것은 무엇인가?

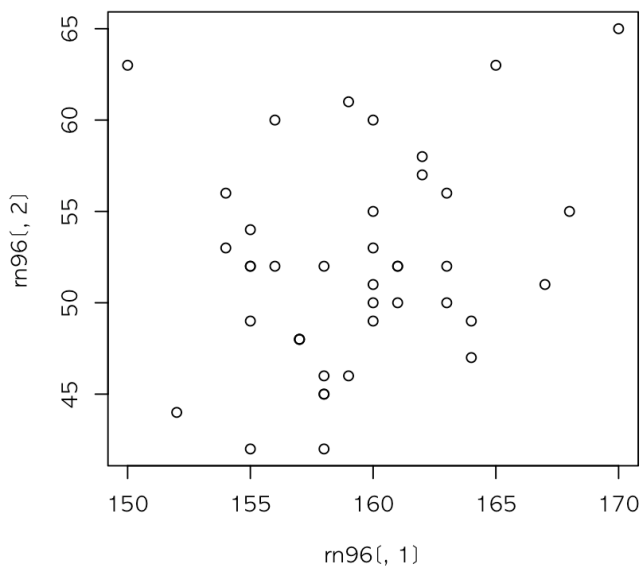
가.



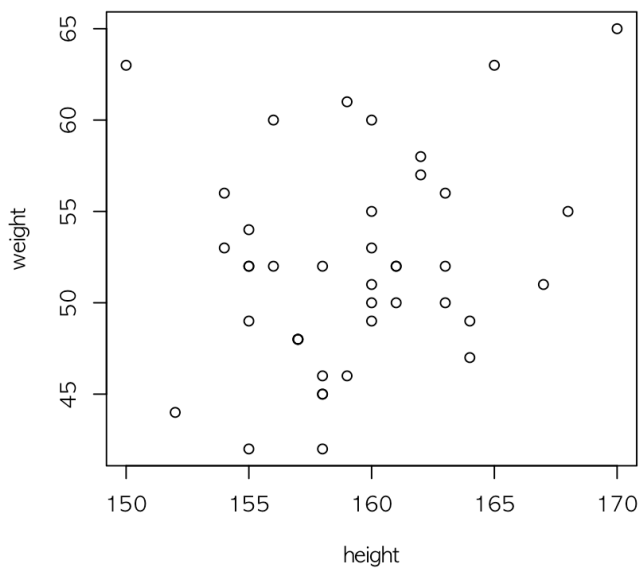
나.



다.

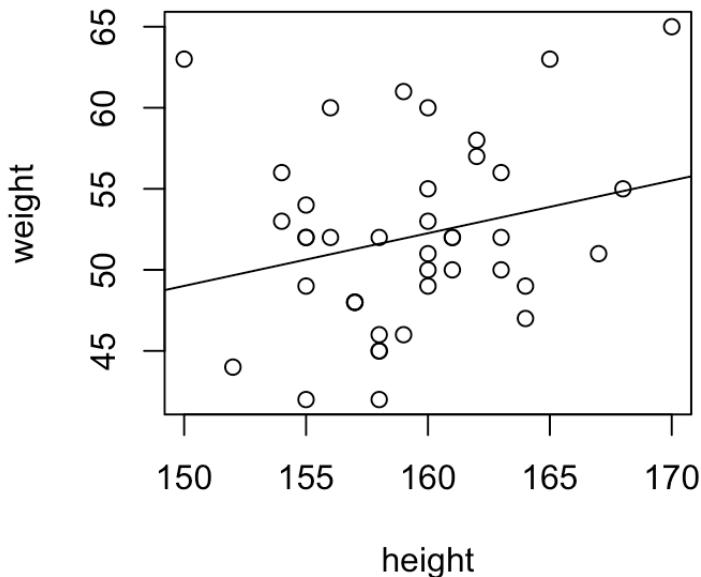


라.



2. 다음 산점도에 회귀선을 그리는 데 적합한 것은 무엇인가?

- 가. `lines(lm(weight ~ height, data = rn96)$coefficient)`
- 나. `points(lm(weight ~ height, data = rn96)$coefficient)`
- 다. `abline(lm(weight ~ height, data = rn96)$coefficient)`
- 라. `polygon(lm(weight ~ height, data = rn96)$coefficient)`



3. 1차 회귀식으로는 살피기 힘든 국소적인 변화를 살피기 위하여 붉은 색으로 local smoother를 추가한다. 이 때 적합한 것은?

- 가. `lines(lowess(rn96$height, rn96$weight), col = "red")`
- 나. `points(lowess(rn96$height, rn96$weight), col = "red")`
- 다. `abline(lowess(rn96$height, rn96$weight), col = "red")`
- 라. `polygon(lowess(rn96$height, rn96$weight), col = "red")`

4. 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 적당한 R 함수는 무엇인가?

- 가. `ifelse`
- 나. `if`
- 다. `for`
- 라. `while`

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

Quetelet's Scottish Soldiers

5. 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 = 480)`

라. ``

| MESURES de la POITRINE. | NOMBRE d'hommes. | NOMBRE PROPORTIONNEL. | PROBABILITÉ d'après L'OBSERVATION. | RANG dans LA TABLE. | RANG d'après le CALCUL. | PROBABILITÉ d'après LA TABLE. | NOMBRE d'OBSERVATIONS calculé. |
|-------------------------------|---------------------|--------------------------|--|---------------------------|-------------------------------|-------------------------------------|--------------------------------------|
| Pouces. | | | | | | | |
| 33 | 3 | 3 | 0,5000 | | | 0,5000 | 7 |
| 34 | 18 | 31 | 0,4995 | 52 | 50 | 0,4995 | 29 |
| 35 | 81 | 141 | 0,4964 | 42,5 | 42,5 | 0,4964 | 110 |
| 36 | 185 | 322 | 0,4825 | 33,5 | 34,5 | 0,4854 | 323 |
| 37 | 420 | 732 | 0,4501 | 26,0 | 26,5 | 0,4531 | 732 |
| 38 | 740 | 1305 | 0,3769 | 18,0 | 18,5 | 0,3799 | 1333 |
| 39 | 1075 | 1867 | 0,2464 | 10,5 | 10,5 | 0,2466 | 1838 |
| | | | 0,0597 | 2,5 | 2,5 | 0,0628 | |
| 40 | 1079 | 1882 | 0,1285 | 5,5 | 5,5 | 0,1359 | 1987 |
| 41 | 934 | 1628 | 0,2913 | 13 | 13,5 | 0,3054 | 1675 |
| 42 | 658 | 1148 | 0,4001 | 21 | 21,5 | 0,4130 | 1090 |
| 43 | 370 | 645 | 0,4706 | 30 | 29,5 | 0,4690 | 560 |
| 44 | 92 | 160 | 0,4866 | 35 | 37,5 | 0,4911 | 221 |
| 45 | 50 | 87 | 0,4955 | 41 | 45,5 | 0,4980 | 69 |
| 46 | 21 | 38 | 0,4991 | 49,5 | 55,5 | 0,4996 | 16 |
| 47 | 4 | 7 | 0,4998 | 50 | 61,8 | 0,4999 | 3 |
| 48 | 1 | 2 | 0,5000 | | | 0,5000 | 1 |
| | 5758 | 1,0000 | | | | | 1,0000 |

6. 아래와 같은 구조를 가진 `chest.table` 에서 "Freq" 를 추출하는 방법 중 그 결과가 다른 것은?

가. `chest.table$Freq`

나. `chest.table[, 2]`

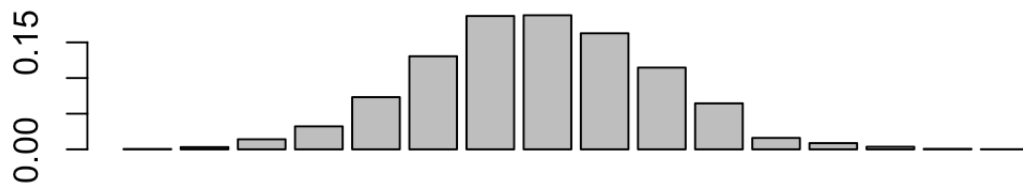
다. `chest.table[, "Freq"]`

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 ...
```

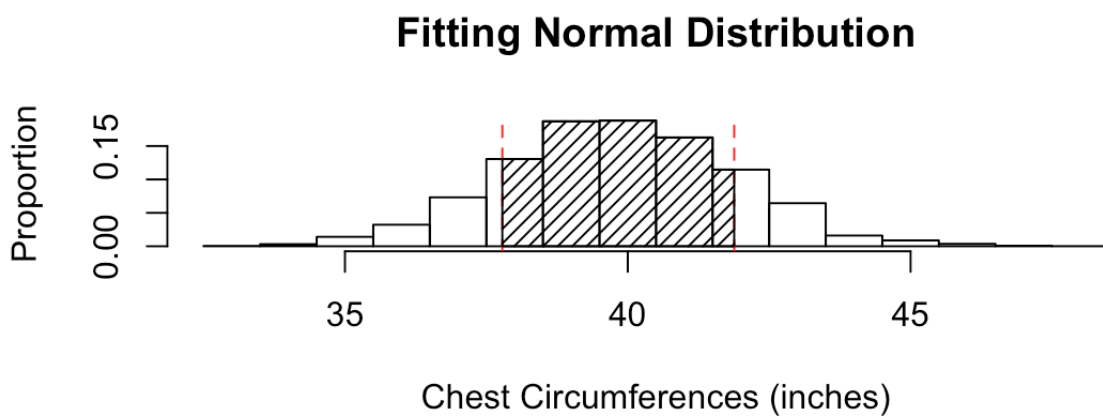
7. 다음 `barplot` 에서 막대 사이의 간격을 없애려면 어떤 조건을 넣어야 하는가?

- 가. `names.arg = 33:48`
- 나. `space = 0`
- 다. `offset = 0`
- 라. `beside = TRUE`



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

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



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

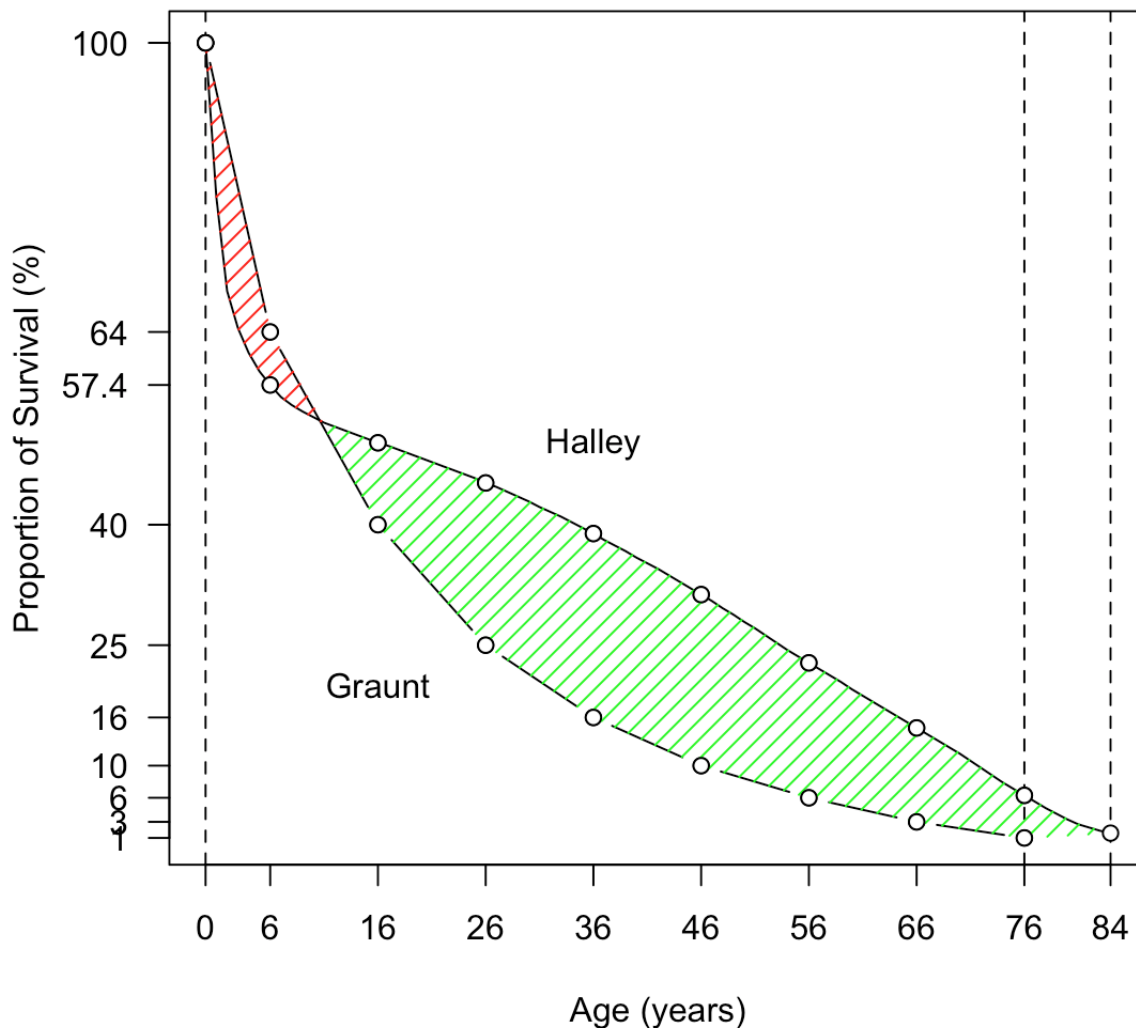
- 가. `rep(chest.table$Chest, chest.table$Freq)`
- 나. `rep(chest.table$Freq, chest.table$Chest)`
- 다. `rep(chest.table$Chest, times = chest.table$Freq)`
- 라. `rep(chest.table$Freq, times = chest.table$Chest)`

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



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

가. `ann = FALSE`

나. `xaxt = "n"`

다. `yaxt = "n"`

라. `type = "l"`

11. 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)`

12. 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)`

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

가. `angle = 45`

나. `density = 15`

다. `col = "green"`

라. `border = NA`

14. 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")`

15. 아래의 reshaping 과정에서 factor를 설정하는 부분은 어디인가?

가. `list(graunt.2, halley.2, us93.2)`

나. `id.vars = "x"`

다. `value.name = "xPo"`

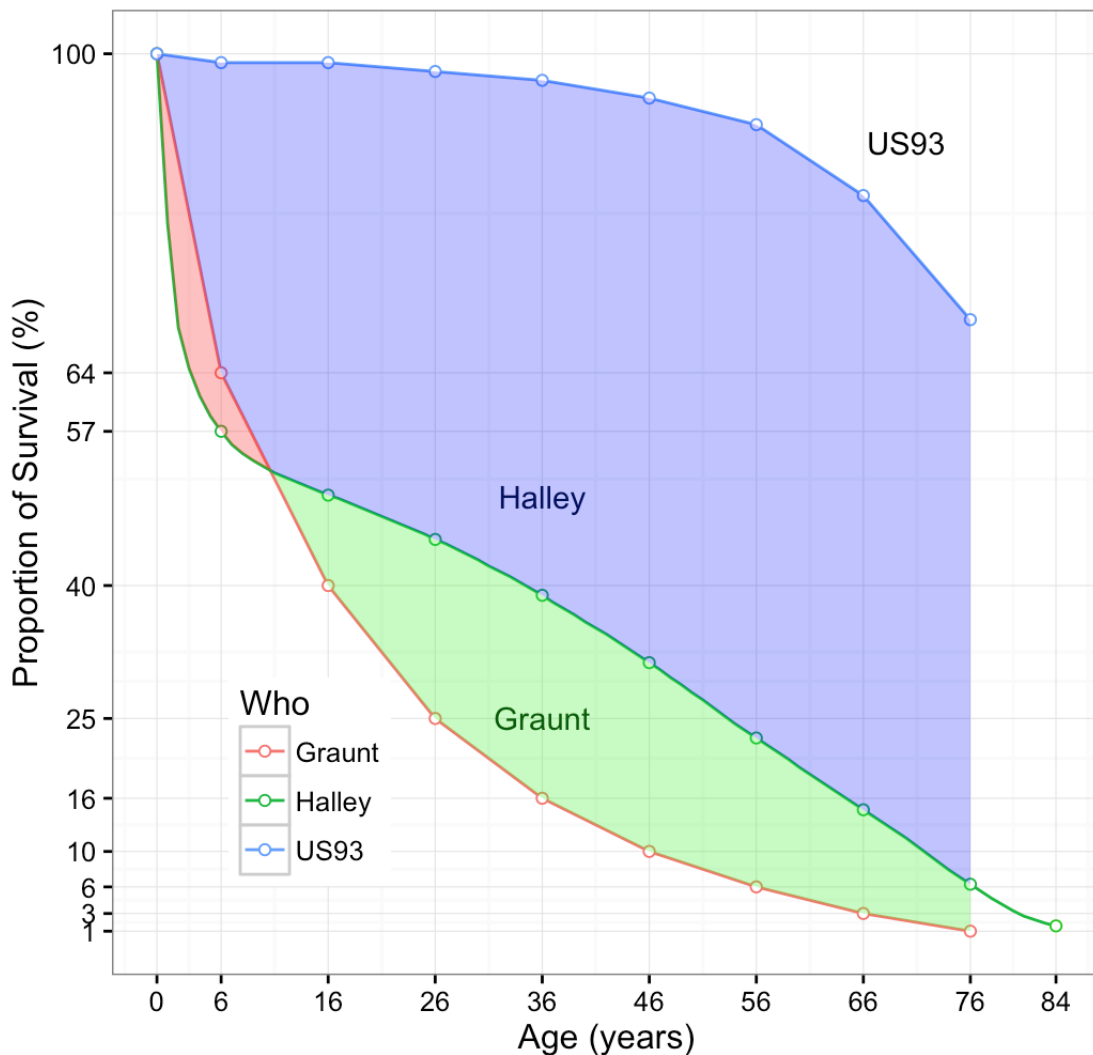
라. `variable.name = "Who"`

```
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", "H
alley", "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 = "gr
een") +
  geom_polygon(data = poly.us.76, aes(x = x, y = y), alpha = 0.3, fill = "blu
e") +
  geom_point(data = data.frame(x = 84, y = halley$xPo[85]), aes(x = x, y = y),
    colour = 3, shape = 21, fill = "white"))
```

Survival Function Plots



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

- 가. `x = x, y = xPo`
- 나. `colour = Who`
- 다. `shape = 21`
- 라. `fill = "white"`

17. 도표의 배경을 흑백으로 바꿔주는 코드는 ?

- 가. `theme_bw()`
- 나. `xlab(x.lab)`
- 다. `ylab(y.lab)`
- 라. `ggtitle(main.title.3)`

18. 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))`

19. 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))`

20. 색깔 투명도를 조절하는 코드는 무엇인가?

- 가. `alpha = 0.3`
- 나. `colour = Who`
- 다. `shape = 21`
- 라. `fill = "white"`