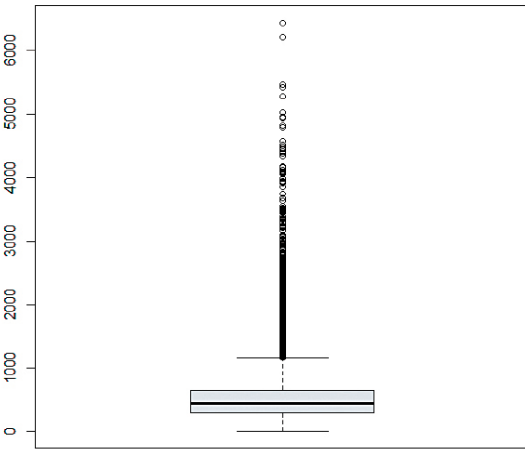
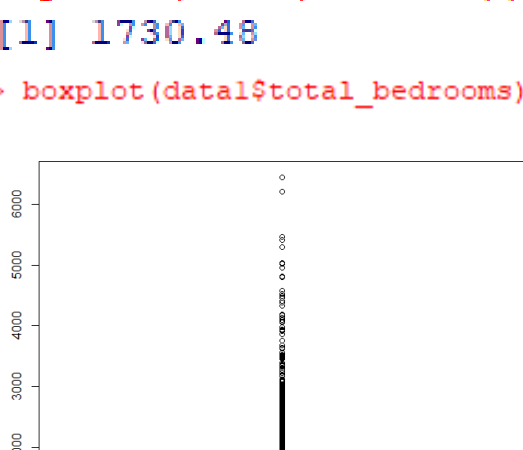


위치	오류유형	수정 전	수정 후
450p	해설	정답 : 5.1 해설 : subset()으로 품종이 setosa인 붓꽃 데이터를 저장 (data, 50개)한다. ifelse() 함수를 이용하여 꽃받침 길이가 중앙값보다 큰 경우 1, 아니면 0의 값을 갖는 새로운 항목 (data\$value)을 추가한다. mean()을 이용하여 평균 (5.1)을 출력하고, 이 값은 모든 데이터에 대한 평균값 (5.006)과 비교하여 높은 값을 가진다는 것을 알 수 있다. <pre> > data <- subset(iris, Species == "setosa") > head(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species 1 5.1 3.5 1.4 0.2 setosa 2 4.9 3.0 1.4 0.2 setosa 3 4.7 3.2 1.3 0.2 setosa 4 4.6 3.1 1.5 0.2 setosa 5 5.0 3.6 1.4 0.2 setosa 6 5.4 3.9 1.7 0.4 setosa > dim(data) [1] 50 5 > summary(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species Min. :4.300 Min. :2.300 Min. :1.000 Min. :0.100 setosa :50 1st Qu.:4.800 1st Qu.:3.200 1st Qu.:1.400 1st Qu.:0.200 versicolor: 0 Median :5.000 Median :3.400 Median :1.500 Median :0.200 virginica : 0 Mean :5.006 Mean :3.428 Mean :1.462 Mean :0.246 3rd Qu.:5.200 3rd Qu.:3.675 3rd Qu.:1.575 3rd Qu.:0.300 Max. :5.800 Max. :4.400 Max. :1.900 Max. :0.600 > > data\$value <- ifelse(data\$Sepal.Length > median(data\$Sepal.Length), 1, 0) > head(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species value 1 5.1 3.5 1.4 0.2 setosa 1 2 4.9 3.0 1.4 0.2 setosa 0 3 4.7 3.2 1.3 0.2 setosa 0 4 4.6 3.1 1.5 0.2 setosa 0 5 5.0 3.6 1.4 0.2 setosa 0 6 5.4 3.9 1.7 0.4 setosa 1 > sum(data\$value) [1] 22 > > mean(data\$Sepal.Length[data\$value]) [1] 5.1 > > summary(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species value Min. :4.300 Min. :2.300 Min. :1.000 Min. :0.100 setosa :50 Min. :0.000 1st Qu.:4.800 1st Qu.:3.200 1st Qu.:1.400 1st Qu.:0.200 versicolor: 0 1st Qu.:0.000 Median :5.000 Median :3.400 Median :1.500 Median :0.200 virginica : 0 Median :0.000 Mean :5.006 Mean :3.428 Mean :1.462 Mean :0.246 Mean :0.44 3rd Qu.:5.200 3rd Qu.:3.675 3rd Qu.:1.575 3rd Qu.:0.300 3rd Qu.:1.000 Max. :5.800 Max. :4.400 Max. :1.900 Max. :0.600 Max. :1.000 </pre>	정답 : 5.313636 해설 : subset()으로 품종이 setosa인 붓꽃 데이터를 저장 (data, 50개)한다. ifelse() 함수를 이용하여 꽃받침 길이가 중앙값보다 큰 경우 1, 아니면 0의 값을 갖는 새로운 항목 (data\$value)를 추가(data\$value는 numeric 변수)한다. mean(data\$Sepal.Length[data\$value == 1])을 이용하여 평균 (5.313636)을 출력하고, 이 값은 모든 데이터에 대한 평균값 (5.006)과 비교하여 높은 값을 가진다는 것을 알 수 있다. 한편, ifelse() 함수 이용 시 꽃받침 길이가 중앙값보다 큰 경우 TRUE, 아니면 FALSE의 값을 갖는 경우로 지정 (data\$value는 logical 변수)하는 경우 평균은 mean(data\$Sepal.Length[data\$value])으로 지정하여 구할 수 있다. <pre> > data <- subset(iris, Species == "setosa") > head(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species 1 5.1 3.5 1.4 0.2 setosa 2 4.9 3.0 1.4 0.2 setosa 3 4.7 3.2 1.3 0.2 setosa 4 4.6 3.1 1.5 0.2 setosa 5 5.0 3.6 1.4 0.2 setosa 6 5.4 3.9 1.7 0.4 setosa > dim(data) [1] 50 5 > summary(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species Min. :4.300 Min. :2.300 Min. :1.000 Min. :0.100 setosa :50 1st Qu.:4.800 1st Qu.:3.200 1st Qu.:1.400 1st Qu.:0.200 versicolor: 0 Median :5.000 Median :3.400 Median :1.500 Median :0.200 virginica : 0 Mean :5.006 Mean :3.428 Mean :1.462 Mean :0.246 3rd Qu.:5.200 3rd Qu.:3.675 3rd Qu.:1.575 3rd Qu.:0.300 Max. :5.800 Max. :4.400 Max. :1.900 Max. :0.600 > > data\$value <- ifelse(data\$Sepal.Length > median(data\$Sepal.Length), 1, 0) > head(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species value 1 5.1 3.5 1.4 0.2 setosa 1 2 4.9 3.0 1.4 0.2 setosa 0 3 4.7 3.2 1.3 0.2 setosa 0 4 4.6 3.1 1.5 0.2 setosa 0 5 5.0 3.6 1.4 0.2 setosa 0 6 5.4 3.9 1.7 0.4 setosa 1 > sum(data\$value) [1] 22 > > class(data\$value) [1] "numeric" > mean(data\$Sepal.Length[data\$value==1]) [1] 5.313636 > > data\$value <- ifelse(data\$Sepal.Length > median(data\$Sepal.Length), TRUE, FALSE) > class(data\$value) [1] "logical" > > mean(data\$Sepal.Length[data\$value]) [1] 5.313636 > summary(data) Sepal.Length Sepal.Width Petal.Length Petal.Width Species value Min. :4.300 Min. :2.300 Min. :1.000 Min. :0.100 setosa :50 Mode :logical 1st Qu.:4.800 1st Qu.:3.200 1st Qu.:1.400 1st Qu.:0.200 versicolor: 0 FALSE:28 Median :5.000 Median :3.400 Median :1.500 Median :0.200 virginica : 0 TRUE :22 Mean :5.006 Mean :3.428 Mean :1.462 Mean :0.246 3rd Qu.:5.200 3rd Qu.:3.675 3rd Qu.:1.575 3rd Qu.:0.300 Max. :5.800 Max. :4.400 Max. :1.900 Max. :0.600 </pre>

위치	오류유형	수정 전	수정 후
457~458p	해설	정답 : -10.08403 해설 : order() 함수를 이용하여 Solar.R 항목을 내림차순 정렬하고 80%의 데이터를 저장한다. 평균값으로 대체하기 전 중앙값(median_before = 39)을 구하고 평균(49.16807)을 결측값으로 대체(ifelse() 함수 이용)한 후 중앙값(49.08403)을 구한다. (평균값 대체 전 중앙값) - (평균값 대체 후 중앙값) = 39 - 49.08403 = -10.08403이다. <pre> > data <- airquality[order(-airquality\$Solar.R),] > head(data) Ozone Solar.R Wind Temp Month Day 16 14 334 11.5 64 5 16 45 NA 332 13.8 80 6 14 41 39 323 11.5 87 6 10 19 30 322 11.5 68 5 19 46 NA 322 11.5 79 6 15 22 11 320 16.6 73 5 22 > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00 Min. :5.000 1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00 1st Qu.:6.000 Median : 31.50 Median :205.0 Median : 9.700 Median :79.00 Median :7.000 Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88 Mean :6.993 3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :37 NA's :7 Day Min. : 1.0 1st Qu.: 8.0 Median :16.0 Mean :15.8 3rd Qu.:23.0 Max. :31.0 > data <- data[!is.na(data\$Ozone),] > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 6.00 Min. : 78.0 Min. : 2.300 Min. :57.00 Min. :5.000 1st Qu.: 22.00 1st Qu.:167.0 1st Qu.: 7.400 1st Qu.:75.00 1st Qu.:6.000 Median : 39.00 Median :224.5 Median : 9.700 Median :81.00 Median :7.000 Mean : 49.17 Mean :214.7 Mean : 9.676 Mean :79.75 Mean :7.059 3rd Qu.: 73.00 3rd Qu.:264.0 3rd Qu.:11.500 3rd Qu.:86.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :33 Day Min. : 1.00 1st Qu.: 7.00 Median :15.00 Mean :15.25 3rd Qu.:24.00 Max. :31.00 > median_before <- median(data\$Ozone, na.rm = TRUE) > median_before [1] 39 > mean <- mean(data\$Ozone, na.rm = TRUE) > mean [1] 49.16807 > data\$Ozone <- ifelse(is.na(data\$Ozone), mean, data\$Ozone) > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 6.00 Min. : 78.0 Min. : 2.300 Min. :57.00 Min. :5.000 1st Qu.: 27.75 1st Qu.:167.0 1st Qu.: 7.400 1st Qu.:75.00 1st Qu.:6.000 Median : 49.08 Median :224.5 Median : 9.700 Median :81.00 Median :7.000 Mean : 49.17 Mean :214.7 Mean : 9.676 Mean :79.75 Mean :7.059 3rd Qu.: 61.50 3rd Qu.:264.0 3rd Qu.:11.500 3rd Qu.:86.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :33 > median_after <- median(data\$Ozone) > median_after [1] 49.08403 > print(median_before - median_after) [1] -10.08403 </pre>	정답 : -10.65217 해설 : order() 함수를 이용하여 Solar.R 항목을 내림차순 정렬하고 80%의 데이터를 저장한다. 평균값으로 대체하기 전 중앙값(median_before = 37)을 구하고 평균(47.65217)을 결측값으로 대체(ifelse() 함수 이용)한 후 중앙값(47.65217)을 구한다. (평균값 대체 전 중앙값) - (평균값 대체 후 중앙값) = 37 - 47.65217 = -10.65217이다. <pre> > data <- airquality[order(-airquality\$Solar.R),] > head(data) Ozone Solar.R Wind Temp Month Day 16 14 334 11.5 64 5 16 45 NA 332 13.8 80 6 14 41 39 323 11.5 87 6 10 19 30 322 11.5 68 5 19 46 NA 322 11.5 79 6 15 22 11 320 16.6 73 5 22 > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 1.00 Min. : 7.0 Min. : 1.700 Min. :56.00 Min. :5.000 1st Qu.: 18.00 1st Qu.:115.8 1st Qu.: 7.400 1st Qu.:72.00 1st Qu.:6.000 Median : 31.50 Median :205.0 Median : 9.700 Median :79.00 Median :7.000 Mean : 42.13 Mean :185.9 Mean : 9.958 Mean :77.88 Mean :6.993 3rd Qu.: 63.25 3rd Qu.:258.8 3rd Qu.:11.500 3rd Qu.:85.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :37 NA's :7 Day Min. : 1.0 1st Qu.: 8.0 Median :16.0 Mean :15.8 3rd Qu.:23.0 Max. :31.0 > data <- data[!is.na(data\$Ozone),] > dim(data) [1] 122 6 > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 6.00 Min. : 78.0 Min. : 2.300 Min. :57.00 Min. :5.000 1st Qu.: 21.00 1st Qu.:165.0 1st Qu.: 7.400 1st Qu.:75.00 1st Qu.:6.000 Median : 37.00 Median :224.5 Median : 9.700 Median :81.00 Median :7.000 Mean : 47.65 Mean :214.9 Mean : 9.765 Mean :79.63 Mean :7.049 3rd Qu.: 71.50 3rd Qu.:264.0 3rd Qu.:11.500 3rd Qu.:86.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :30 > median_before <- median(data\$Ozone, na.rm = TRUE) > median_before [1] 37 > mean <- mean(data\$Ozone, na.rm = TRUE) > mean [1] 47.65217 > data\$Ozone <- ifelse(is.na(data\$Ozone), mean, data\$Ozone) > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 6.00 Min. : 78.0 Min. : 2.300 Min. :57.00 Min. :5.000 1st Qu.: 24.75 1st Qu.:165.0 1st Qu.: 7.400 1st Qu.:75.00 1st Qu.:6.000 Median : 47.65 Median :224.5 Median : 9.700 Median :81.00 Median :7.000 Mean : 47.65 Mean :214.9 Mean : 9.765 Mean :79.63 Mean :7.049 3rd Qu.: 81.50 3rd Qu.:264.0 3rd Qu.:11.500 3rd Qu.:86.00 3rd Qu.:8.000 Max. :168.00 Max. :334.0 Max. :20.700 Max. :97.00 Max. :9.000 NA's :31 > median_after <- median(data\$Ozone) > median_after [1] 47.65217 > print(median_before - median_after) [1] -10.65217 </pre>

위치	오류유형	수정 전	수정 후
459p	해설	정답 : 14. 89663 해설 : na.omit()으로 결측값 제거 후 quantile() 함수를 이용하여 Ozone 항목에 대한 사분위(q)를 구한다. q[2]는 하위 25%의 값(18), q[4]는 상위 25%의 값(62)이며, ifelse()로 해당 값을 만족하는 Ozone 항목을 0으로 대체한다. 대체된 데이터세트를 이용하여 평균 + 표준편차 $= \text{mean}(\text{data}\$Ozone) + \text{sd}(\text{data}\$Ozone) \text{ 값을 출력한다.}$ <pre> > data <- na.omit(airquality) > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 1.0 Min. : 7.0 Min. : 2.30 Min. :57.00 Min. :5.000 Min. : 1.00 1st Qu.:18.0 1st Qu.:113.5 1st Qu.: 7.40 1st Qu.:71.00 1st Qu.:6.000 1st Qu.: 9.00 Median :31.0 Median :207.0 Median : 9.70 Median :79.00 Median :7.000 Median :16.00 Mean :42.1 Mean :184.8 Mean : 9.94 Mean :77.79 Mean :7.216 Mean :15.95 3rd Qu.:62.0 3rd Qu.:255.5 3rd Qu.:11.50 3rd Qu.:84.50 3rd Qu.:9.000 3rd Qu.:22.50 Max. :168.0 Max. :334.0 Max. :20.70 Max. :97.00 Max. :9.000 Max. :31.00 > > q <- quantile(data\$Ozone) > q 0% 25% 50% 75% 100% 18 31 62 168 > str(q) Named num [1:5] 1 18 31 62 168 - attr(*, "names")= chr [1:5] "0%" "25%" "50%" "75%" ... > > data\$Ozone <- ifelse(data\$Ozone >= q[3] + data\$Ozone <= q[2], + 0, data\$Ozone) > > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 0.000 Min. : 7.0 Min. : 2.30 Min. :57.00 Min. :5.000 Min. : 1.00 1st Qu.: 0.000 1st Qu.:113.5 1st Qu.: 7.40 1st Qu.:71.00 1st Qu.:6.000 1st Qu.: 9.00 Median : 0.000 Median :207.0 Median : 9.70 Median :79.00 Median :7.000 Median :16.00 Mean : 5.072 Mean :184.8 Mean : 9.94 Mean :77.79 Mean :7.216 Mean :15.95 3rd Qu.: 0.000 3rd Qu.:255.5 3rd Qu.:11.50 3rd Qu.:84.50 3rd Qu.:9.000 3rd Qu.:22.50 Max. :30.000 Max. :334.0 Max. :20.70 Max. :97.00 Max. :9.000 Max. :31.00 > > print(mean(data\$Ozone) + sd(data\$Ozone)) [1] 14.89663 </pre>	정답 : 34. 53803 해설 : na.omit()으로 결측값 제거 후, quantile() 함수를 이용하여 Ozone 항목에 대한 사분위(q)를 구한다. q[2]는 하위 25%의 값(18), q[4]는 상위 25%의 값(62)이며, ifelse()로 해당 값을 만족하는 Ozone 항목을 0으로 대체한다. 대체된 데이터세트를 이용하여 평균 + 표준편차 = $\text{mean}(\text{data}\$Ozone) + \text{sd}(\text{data}\$Ozone) \text{ 값을 출력(34. 53803) 한다.}$ <pre> > data <- na.omit(airquality) > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 1.0 Min. : 7.0 Min. : 2.30 Min. :57.00 Min. :5.000 Min. : 1.00 1st Qu.:18.0 1st Qu.:113.5 1st Qu.: 7.40 1st Qu.:71.00 1st Qu.:6.000 1st Qu.: 9.00 Median :31.0 Median :207.0 Median : 9.70 Median :79.00 Median :7.000 Median :16.00 Mean :42.1 Mean :184.8 Mean : 9.94 Mean :77.79 Mean :7.216 Mean :15.95 3rd Qu.:62.0 3rd Qu.:255.5 3rd Qu.:11.50 3rd Qu.:84.50 3rd Qu.:9.000 3rd Qu.:22.50 Max. :168.0 Max. :334.0 Max. :20.70 Max. :97.00 Max. :9.000 Max. :31.00 > > q <- quantile(data\$Ozone) > q 0% 25% 50% 75% 100% 18 31 62 168 > > str(q) Named num [1:5] 1 18 31 62 168 - attr(*, "names")= chr [1:5] "0%" "25%" "50%" "75%" ... > > data\$Ozone <- ifelse(data\$Ozone >= q[4] + data\$Ozone <= q[2], + 0, data\$Ozone) > > summary(data) Ozone Solar.R Wind Temp Month Day Min. : 0.00 Min. : 7.0 Min. : 2.30 Min. :57.00 Min. :5.000 Min. : 1.00 1st Qu.: 0.00 1st Qu.:113.5 1st Qu.: 7.40 1st Qu.:71.00 1st Qu.:6.000 1st Qu.: 9.00 Median : 0.00 Median :207.0 Median : 9.70 Median :79.00 Median :7.000 Median :16.00 Mean :15.83 Mean :184.8 Mean : 9.94 Mean :77.79 Mean :7.216 Mean :15.95 3rd Qu.:30.50 3rd Qu.:255.5 3rd Qu.:11.50 3rd Qu.:84.50 3rd Qu.:9.000 3rd Qu.:22.50 Max. :61.00 Max. :334.0 Max. :20.70 Max. :97.00 Max. :9.000 Max. :31.00 > > print(mean(data\$Ozone) + sd(data\$Ozone)) [1] 34.53803 </pre>

위치	오류유형	수정 전	수정 후
481p	문제-본문	정답 : 이상치들의 평균 =1,795.81 해설 : data1의 total_bedrooms 항목의 평균(m)과 표준편차(n)를 구하고 이상값 판별을 위한 하한값(Low)과 상한값(Upper)을 정의한다. 상한값 이상 그리고 하한값 이하인 값들을 이상치로 정의(result)하여 평균(mean(outlier))을 출력한다. boxplot() 함수를 이용하여 항목 값들의 이상값을 개략적으로 확인한다. <pre>> m <- mean(data\$total_bedrooms) > m [1] 543.2013 > n <- sd(data\$total_bedrooms) > n [1] 433.9254 > > Low <- m - n*1.5 > Low [1] -107.6868 > Upper <- m + n*1.5 > Upper [1] 1194.089 > > result <- data\$total_bedrooms >= Upper data\$total_bedrooms <= Low > result [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [16] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [31] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [46] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [76] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE > > outlier <- data\$total_bedrooms[result] > outlier [1] 2477 1331 1270 1414 1603 1914 1196 1750 1344 2048 1212 1744 2408 1249 2885 1379 1554 1270 [19] 2045 1309 1526 1326 1279 1207 1439 2031 1253 1516 1374 1273 2993 2708 1407 1376 1818 2861 [37] 1480 1294 1823 1247 1375 2090 1540 1207 1249 3864 1588 1373 1390 1576 1207 1294 2074 1344 [55] 1510 2220 3493 1217 2210 1821 1578 1785 1860 3288 1375 1355 1559 1474 1288 1204 1522 1314 [73] 2252 1646 1284 1404 1882 1639 1534 1691 1939 1213 1603 2558 1269 1422 1685 1590 1521 1455 [91] 1487 2826 1439 1364 1248 1424 1359 1556 1276 2546 1355 1411 2275 1426 1335 1717 1382 1739 [109] 1578 1250 1210 1551 1670 1384 1201 1200 1200 1439 1330 1346 2401 1429 1486 2244 2121 2190 [127] 2139 2685 1489 2141 1901 1455 1994 1369 1517 1597 1453 1600 1257 1482 1692 1657 1492 1636 [145] 1514 1653 1490 1286 1443 2139 1464 1872 1653 1506 2355 1325 2387 1209 1464 1537 2717 1459 [163] 1768 2193 1767 1369 2814 1209 1293 1527 1283 1611 2007 1271 1477 1214 1594 1477 1350 1820 [181] 1358 1737 1826 1849 1304 1317 2560 1869 4183 2691 1345 1412 1994 1976 4457 1995 1295 1706 > > print(mean(outlier)) [1] 1795.81 > boxplot(data\$total_bedrooms)</pre> 	정답 : 이상치들의 평균 = 1730.48 해설 : 결측치를 제외한 total_bedrooms의 중앙값(median) = 435이다. 결측값을 중앙값으로 대체한 후 이를 data1에 저장한다. data1의 total_bedrooms 항목의 평균(m = 536.8389)과 표준편차(n = 419.3919)를 구하고 이상값 판별을 위한 하한값(Low = -92.24896)과 상한값(Upper = 1165.927)을 정의한다. 상한값 이상 그리고 하한값 이하인 값들을 이상치로 정의(result)하여 평균(mean(outlier) = 1730.48)을 출력한다. boxplot() 함수를 이용하여 항목 값들의 이상값을 개략적으로 확인한다. <pre>> median <- median(data\$total_bedrooms, na.rm=TRUE) > median [1] 435 > data\$total_bedrooms <- ifelse(is.na(data\$total_bedrooms), median, data\$total_bedrooms) > data1 <- data > summary(data1) longitude latitude housing_median_age total_rooms total_bedrooms population households median_income 1st Qu.:121.4 1st Qu.:121.93 1st Qu.:118.00 1st Qu.:1440 1st Qu.:237.0 1st Qu.:787 1st Qu.:250.0 1st Qu.:2.5624 Mean:112.4 Median:121.54 Mean:127.00 Mean:15237 Mean:536.8 Mean:1165.9 Mean:409.0 Mean:3.7481 Max.:114.0 Max.:141.90 Max.:152.00 Max.:19320 Max.:4495.0 Max.:15652 Max.:1602.0 Max.:15.0001 Median:104.9 Mean:126.60 Clear:104.90 Median:1179700 Mode:104.90 Max.:1248725 Max.:1600001 > m <- mean(data\$total_bedrooms) > m [1] 536.8389 > n <- sd(data\$total_bedrooms) > n [1] 419.3919 > > Low <- m - n*1.5 > Low [1] -92.24896 > Upper <- m + n*1.5 > Upper [1] 1165.927 > > result <- data\$total_bedrooms >= Upper data\$total_bedrooms <= Low > result [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [16] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [31] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [46] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [76] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [91] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [106] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [136] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [151] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [166] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [196] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [211] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [226] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [256] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [271] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [286] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [316] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [331] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [346] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [376] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [391] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [406] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [436] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [451] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [466] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [496] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [511] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [526] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [541] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [556] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [571] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [586] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [601] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [616] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [631] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [646] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [661] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [676] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [691] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [706] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [721] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [736] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [751] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [766] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [781] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [796] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [811] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [826] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [841] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [856] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [871] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [886] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [901] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [916] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [931] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [946] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [976] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE [991] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE > > print(mean(outlier)) [1] 1730.48 > boxplot(data\$total_bedrooms)</pre> 

도서의 오류로 학습에 불편드린 점 진심으로 사과드립니다.
더 나은 도서를 만들기 위해 노력하는 시대교육그룹이 되겠습니다.