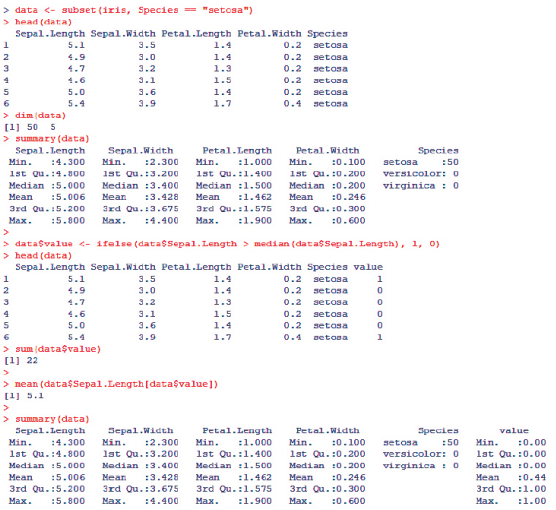
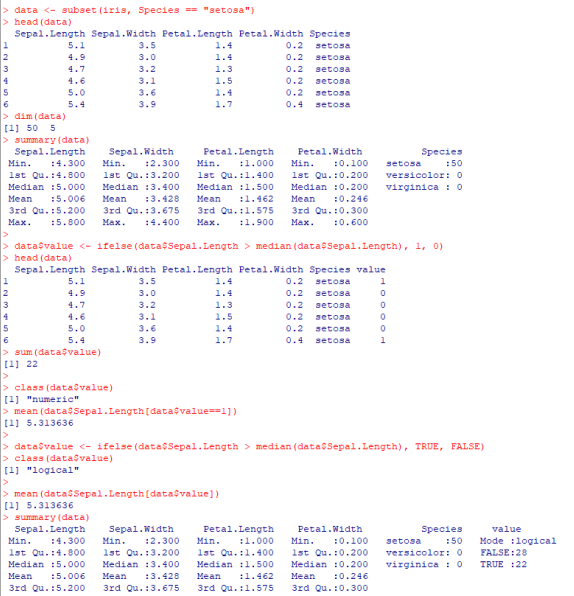
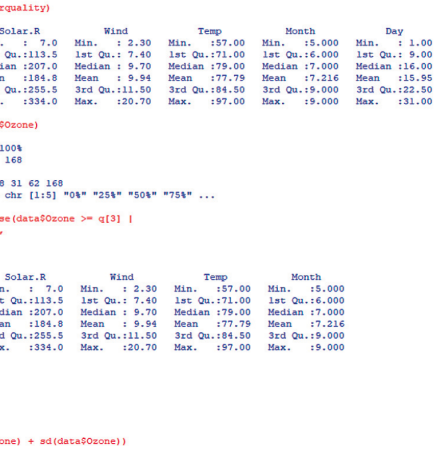
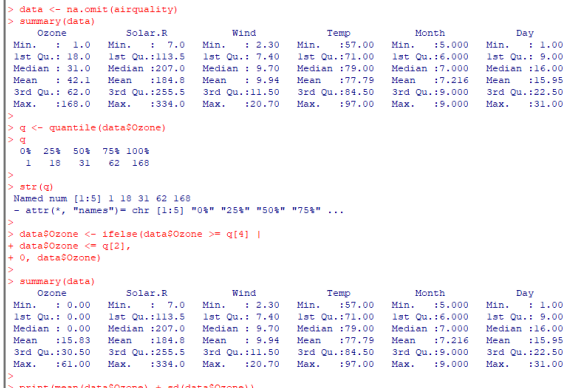
위치	오류유형	수정 전	수정 후
450p	해설	**정답** : 5.1 **해설** : subset( )으로 품종이 setosa인 붓꽃 데이터를 저장(data, 50개)한다. ifelse( ) 함수를 이용하여 꽃받침 길이가 중앙값보다 큰 경우 1, 아니면 0의 값을 갖는 새로운 항목(data$value)을 추가한다. mean( )을 이용하여 평균 (5.1)을 출력하고, 이 값은 모든 데이터에 대한 평균값(5.006)과 비교하여 높은 값을 가진다는 것을 알 수 있다. 	**정답** : 5.313636 **해설** : subset( )으로 품종이 setosa인 붓꽃 데이터를 저장(data, 50개)한다. ifelse( ) 함수를 이용하여 꽃받침 길이가 중앙값보다 큰 경우 1, 아니면 0의 값을 갖는 새로운 항목(data$value)를 추가(data$value는 numeric 변수)한다. mean(data$Sepal. Length[data$value = = 1]) 을 이용하여 평균(5.313636)을 출력하고, 이 값은 모든 데이터에 대한 평균값(5.0006)과 비교하여 높은 값을 가진다는 것을 알 수 있다. 한편, ifelse( ) 함수 이용 시 꽃받침 길이가 중앙값보다 큰 경우 TRUE, 아니면 FALSE의 값을 갖는 경우로 지정(data$value는 logical 변수)하는 경우 평균은 mean(data$Sepal. Length[data$value]) 으로 지정하여 구할 수 있다. 

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.
		`> data <- airquality[order(-airquality$Solar.R),]`
		`> head(data)`
		`> summary(data)`
		`> data <- data[1:nrow(data)*0.8,]`
		`> summary(data)`
		`> median_before <- median(data$Ozone, na.rm = TRUE)`
		`> median_before`
		`> mean <- mean(data$Ozone, na.rm = TRUE)`
		`> mean`
		`> data$Ozone <- ifelse(is.na(data$Ozone), mean, data$Ozone)`
		`> summary(data)`
		`> median_after <- median(data$Ozone)`
		`> median_after`
		`> print(median_before - median_after)`
			정답 : -10.65217
			해설 : order( ) 함수를 이용하여 Solar.R 항목을 내림차순 정렬하고 80%의 데이터를 저장한다. 평균값으로 대체하기 전 중앙값(median_before = 37)을 구하고 평균(47.65217)을 결측값으로 대체(ifelse( ) 함수 이용)한 후 중앙값(47.65217)을 구한다. (평균값 대체 전 중앙값) - (평균값 대체 후 중앙값) = 37 - 47.65217 = -10.65217이다.
			`> data <- airquality[order(-airquality$Solar.R), ]`
			`> head(data)`
			`> summary(data)`
			`> data <- data[1:(nrow(data)*0.8), ]`
			`> dim(data)`
			`> summary(data)`
			`> median_before <- median(data$Ozone, na.rm = TRUE)`
			`> median_before`
			`> mean <- mean(data$Ozone, na.rm = TRUE)`
			`> mean`
			`> data$Ozone <- ifelse(is.na(data$Ozone), mean, data$Ozone)`
			`> summary(data)`
			`> median_after <- median(data$Ozone)`
			`> median_after`
			`> print(median_before - median_after)`

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 = mean(data$Ozone) + sd(data$Ozone) 값을 출력한다. 	**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 = mean(data$Ozone) + sd(data$Ozone) 값을 출력(34.53803)한다. 

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.	**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.

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