**[프라임칼리지] 데이터시각화 - 1차 과제물(100점 만점)**

**제출기한: 2025.04.04.(금) 23:59**

**추가연장 제출기한: 2025.04.05.(토) 23:59**

(‘과제’ 메뉴)

**이 름: 신승엽**

**□ 모든 문항에 대해 답안을 작성하고, ‘R 코드’와 ‘R 프로그래밍을 실행한 결과(캡처 이미지)’를 각각 첨부해주세요. (총 3문항)**

- 모든 문항에 대해 결과에 대한 **해석**까지 반드시 포함해주세요.

- **R 실습 과제**이므로 엑셀이나 공학용 계산기, 파이썬 등 다른 프로그램이나 손으로 직접 계산한 답안은 평가가 불가능합니다.

**문제1.** [5강. ggplot2 II]

R의 내장 데이터셋 중 하나인 'ChickWeight'를 활용하여 시간에 따른 병아리의 체중 변화를 시각화하려고 한다. (30점)

(R에서 ChickWeight를 콘솔에 입력하거나 스크립트에 적어서 실행시키면 불러올 수 있음 – 대소문자 유의)

선행 코드:

# ChickWeight 데이터셋 불러오기

library(ggplot2)

data(ChickWeight)

1. 'Time'을 x축으로, 'weight'를 y축으로 하여 선 그래프를 그리시오. 이때, 각 병아리(Chick) 별 성장 궤적을 구분할 수 있도록 group = Chick을 설정하시오.

코드:

p1 <- ggplot(ChickWeight, aes(x = Time, y = weight, group = Chick)) +

geom\_line() +

labs(title = "병아리 별 체중 변화",

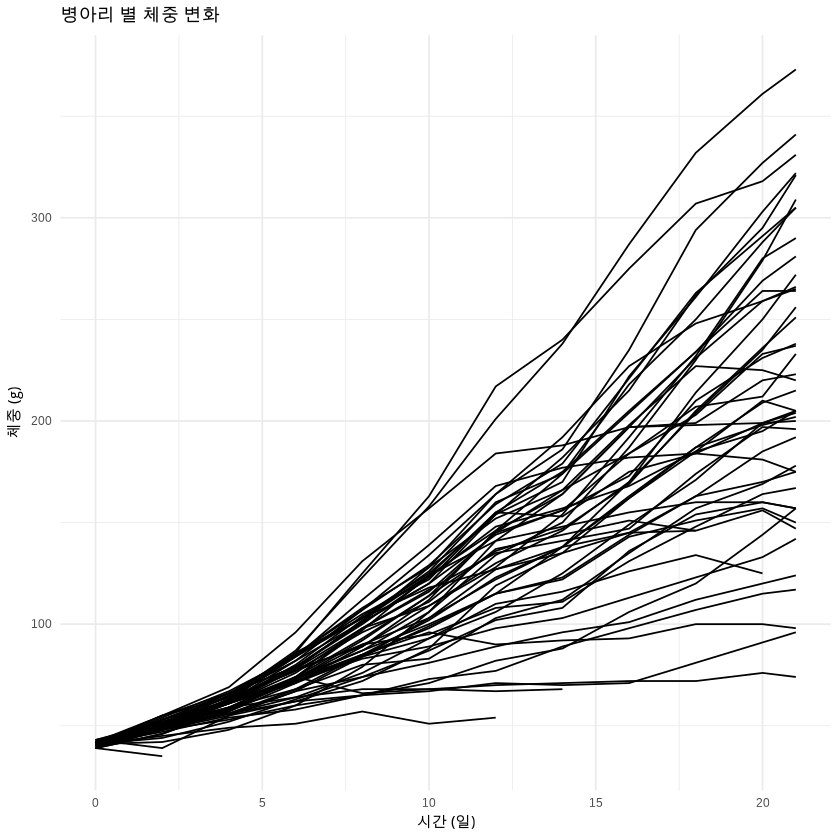
x = "시간 (일)",

y = "체중 (g)") +

theme\_minimal()

print(p1)

결과:



해석:

생성된 그래프(p1)는 ChickWeight 데이터셋에 포함된 각 개별 병아리(Chick)의 시간에 따른 체중(weight) 변화를 선으로 보여줍니다. 시간 축(x축)은 경과 시간을 일 단위로 나타내고, 체중 축(y축)은 병아리의 체중을 그램 단위로 나타냅니다. group = Chick 설정을 통해 각 선은 고유한 병아리 한 마리의 성장 과정을 나타냅니다. 전반적으로 모든 병아리의 체중이 시간이 지남에 따라 증가하는 경향을 보이지만, 개체별로 성장 속도와 최종 도달 체중에 상당한 차이가 있음을 시각적으로 확인할 수 있습니다. 일부 병아리는 다른 병아리보다 훨씬 빠르게 성장하는 반면, 일부는 성장이 더딘 모습을 보입니다.

2. 선의 색상을 병아리의 Diet에 따라 다르게 지정하고, 그래프에 제목과 축 레이블을 적절히 추가하시오.

코드:

p2 <- ggplot(ChickWeight, aes(x = Time, y = weight, group = Chick, color = Diet)) +

geom\_line() +

labs(title = "Diet에 따른 병아리 체중 변화",

x = "시간 (일)",

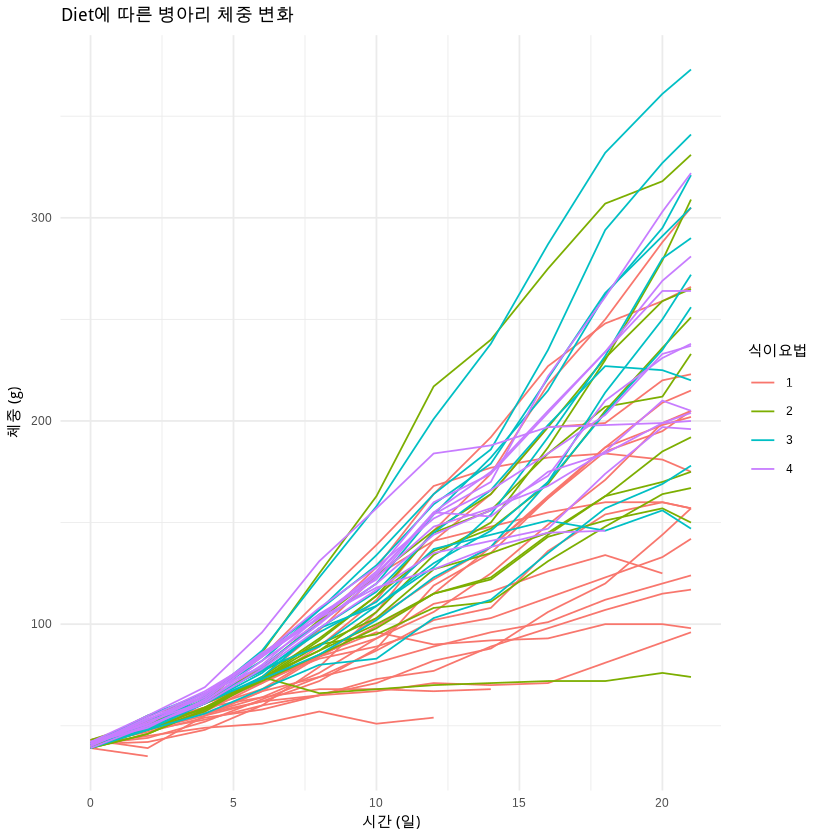
y = "체중 (g)",

color = "식이요법") +

theme\_minimal()

print(p2)

결과:



해석:

생성된 그래프(p2)는 각 개별 병아리의 시간에 따른 체중 변화를 선으로 나타내며, 각 선의 색상은 해당 병아리가 섭취한 Diet(식이요법 1, 2, 3, 4)를 구분합니다. 범례를 통해 어떤 색상이 어떤 Diet를 나타내는지 알 수 있습니다. 그래프에는 "Diet에 따른 병아리 체중 변화"라는 제목과 x축("시간 (일)"), y축("체중 (g)") 레이블이 명확하게 표시되어 있습니다. 이 그래프는 평균적인 추세선을 보여주지 않고, 오직 개별 병아리들의 성장 궤적만을 Diet별로 색상을 달리하여 보여줍니다. 따라서 Diet 그룹 간의 전반적인 성장 경향을 파악하려면 각 색상 그룹에 속한 선들의 전반적인 분포와 최종 도달 위치를 주의 깊게 관찰해야 합니다. 선들이 겹쳐 보일 수 있으나, 색상 구분을 통해 특정 Diet 그룹의 성장 범위나 패턴을 대략적으로 파악할 수 있습니다.

3. 완성된 그래프에 대해 각 Diet별 병아리의 성장 패턴을 비교하여 설명하시오.

- Diet 1 그룹 (특정 색상): 이 색상의 선들은 전체적으로 그래프의 가장 아래쪽에 분포하는 경향을 보입니다. 즉, 시간이 지나도 체중 증가폭이 상대적으로 작으며, 실험 종료 시점(21일)의 체중 값들이 다른 그룹에 비해 낮게 형성되어 있습니다. 이는 Diet 1 그룹의 성장률이 가장 느리다는 것을 시사합니다.

- Diet 2 그룹 (다른 색상): 이 색상의 선들은 Diet 1 그룹의 선들보다는 위쪽에 분포하지만, Diet 3이나 4 그룹보다는 아래쪽에 위치하는 경향이 있습니다. 이는 Diet 2가 Diet 1보다는 높은 성장률을 보이지만, 다른 두 그룹보다는 낮은 중간 정도의 성장 패턴을 가짐을 나타냅니다.

- Diet 3 그룹 (또 다른 색상): 이 색상의 선들은 Diet 2 그룹보다 높은 체중 범위에 분포하며, Diet 4 그룹과 유사한 영역에 많이 퍼져 있습니다. 하지만 전반적인 선들의 최종 도달점을 보면 Diet 4 그룹보다는 약간 낮은 경향을 보일 수 있습니다. 비교적 빠른 성장 패턴을 가집니다.

- Diet 4 그룹 (나머지 색상): 이 색상의 선들이 전반적으로 가장 위쪽에 분포하며, 특히 실험 후반부(Time 값이 클 때)에 높은 체중 값을 갖는 선들이 많습니다. 이는 Diet 4 그룹이 가장 빠른 성장률과 가장 높은 최종 체중을 달성하는 경향이 있음을 보여줍니다.

- 전반적 비교: 각 색상 그룹에 속한 선들의 '묶음'이 y축(체중) 상에서 차지하는 위치를 비교함으로써 Diet별 성장 패턴의 차이를 추론할 수 있습니다. Diet 4가 가장 우수한 성장을, Diet 1이 가장 저조한 성장을 보이며, Diet 2와 3은 그 중간 수준의 성장 패턴을 나타냅니다. 개별 선들의 퍼짐 정도는 각 Diet 그룹 내의 개체 변동성을 보여줍니다.

4. 성장 패턴의 차이가 나타나는 이유에 대해 자신의 견해를 논리적으로 서술하시오.

그래프를 보면 확실히 병아리들이 어떤 사료(Diet)를 먹었는지에 따라 성장하는 모습이 크게 다르다는 것을 알 수 있습니다. 이런 차이가 나타나는 가장 큰 이유는 역시 사료의 영양 구성이 달랐기 때문이라고 생각합니다.

아시다시피 병아리가 쑥쑥 크려면 단백질이나 에너지원, 비타민, 미네랄 같은 필수 영양소들이 충분하고 균형 있게 공급되어야 합니다. 아마 네 가지 Diet는 이런 영양소들의 양이나 비율 면에서 서로 차이가 있었을 것이고, 이것이 곧 성장 속도와 최종 몸무게의 차이로 이어진 것이겠죠.

특히 Diet 4를 먹은 병아리들이 가장 눈에 띄게 잘 자란 것을 보면, 이 사료가 성장에 필요한 영양소들을 아주 효과적인 조합으로 제공했을 가능성이 높습니다. 병아리들이 가진 성장 잠재력을 최대한 끌어내 준 것이 아닐까 싶습니다.

반면에 Diet 1을 먹은 병아리들은 성장이 가장 더뎠는데, 이는 아마 사료에 성장에 필요한 무언가가 부족했거나 영양 균형이 잘 맞지 않았기 때문일 겁니다. 예를 들어 필수 아미노산이 부족했을 수도 있고, 에너지에 비해 단백질이 부족했을 수도 있겠죠.

Diet 2와 3은 그 중간 정도의 성장세를 보였으니, 영양 구성도 딱 중간 수준이었을 것으로 짐작해 볼 수 있습니다.

물론 같은 사료를 먹었더라도 병아리마다 조금씩 다르게 자라는 모습도 보입니다. 이건 병아리 개체가 가진 유전적인 차이나 태어날 때의 건강 상태, 혹은 아주 미세한 사료 섭취량의 차이 같은 다른 요인들도 영향을 주기 때문일 겁니다. 하지만 그룹 전체로 봤을 때 나타나는 뚜렷한 성장 패턴의 차이는 역시 사료, 즉 Diet 자체의 영양학적 특성 때문이라고 설명하는 것이 가장 설득력 있어 보입니다.

결론적으로, 병아리에게 어떤 사료를 먹이느냐는 성장 속도와 최종 체중에 결정적인 영향을 미친다고 할 수 있습니다. 이 실험 결과를 보면, 영양학적으로 잘 설계된 사료(아마도 Diet 4가 여기에 해당하겠죠)가 병아리를 건강하고 빠르게 키우는 데 얼마나 중요한지 다시 한번 확인할 수 있습니다.

**문제2.** [6강. 색(color)]

R의 내장 데이터셋 중 하나인 'mtcars'를 활용하여 자동차의 연비(mpg)와 마력(hp) 간의 관계를 시각화하려고 한다. (30점)

(R에서 mtcars를 콘솔에 입력하거나 스크립트에 적어서 실행시키면 불러올 수 있음)

선행 코드:

# mtcars 데이터셋 불러오기

data(mtcars)

1. x축은 마력(hp), y축은 연비(mpg)로 하고, 실린더 개수(cyl)에 따라 색상을 다르게 표현한 산점도를 그리시오.

코드:

p\_mtcars\_base <- ggplot(mtcars, aes(x = hp, y = mpg, color = as.factor(cyl))) +

geom\_point(size = 3) +

labs(title = "마력(hp)과 연비(mpg)의 관계 (실린더 개수별 색상)",

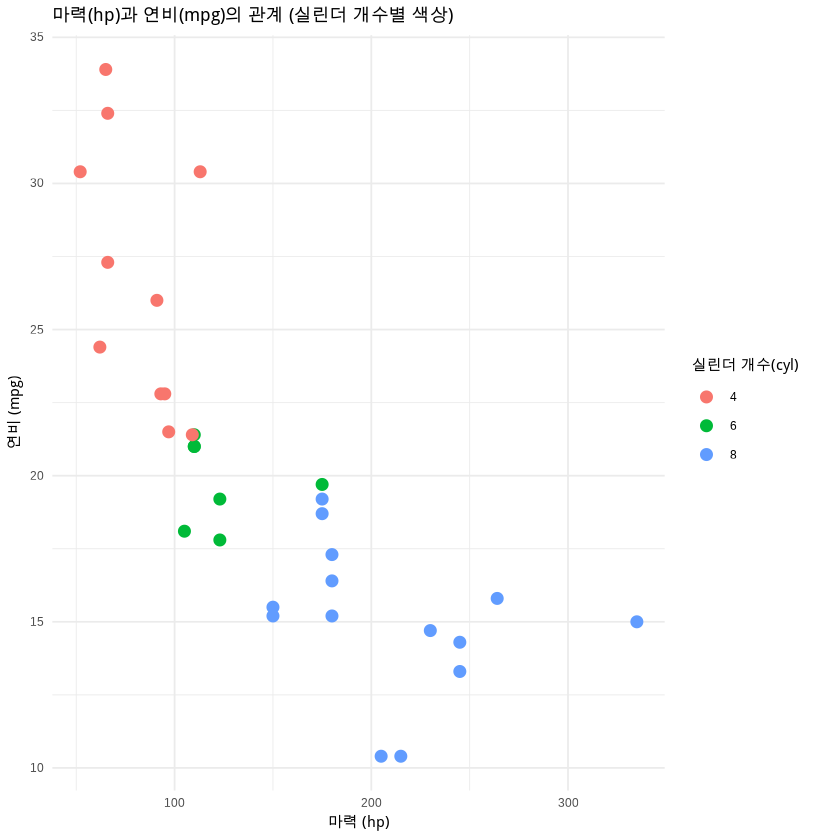
x = "마력 (hp)",

y = "연비 (mpg)",

color = "실린더 개수(cyl)") + theme\_minimal()

print(p\_mtcars\_base)

결과:



해석:

위 코드는 mtcars 데이터셋의 마력(hp)을 x축, 연비(mpg)를 y축으로 하는 산점도를 생성합니다. 각 점은 자동차 모델 하나를 나타냅니다. 점의 색상은 해당 자동차의 실린더 개수(cyl)에 따라 구분됩니다 (예: 4기통, 6기통, 8기통이 각각 다른 색상으로 표시됨). 범례를 통해 어떤 색상이 몇 개의 실린더를 의미하는지 알 수 있습니다. 이 기본 산점도를 통해 마력이 높을수록 연비가 낮아지는 전반적인 음의 상관관계 경향과, 실린더 개수가 다른 자동차들이 마력과 연비 측면에서 어떻게 분포하는지 대략적으로 파악할 수 있습니다.

2. 'viridis' 컬러맵을 사용하여 색상을 지정하고, 생성된 컬러바(color bar)의 방향을 수직에서 수평으로 변경하시오.

코드:

p\_mtcars\_final <- ggplot(mtcars, aes(x = hp, y = mpg, color = cyl)) + # cyl을 숫자로 매핑

geom\_point(size = 3) +

scale\_color\_viridis\_c(option = "viridis", # viridis 연속형 컬러맵 사용

breaks = c(4, 6, 8), # 컬러바에 표시될 값 지정

guide = guide\_colorbar( # 컬러바 가이드 설정

direction = "horizontal", # 방향 수평으로

title.position = "top", # 범례 제목 위치 위쪽

barwidth = unit(10, "lines"), # 컬러바 너비 조절

barheight = unit(0.5, "lines") # 컬러바 높이 조절

)) +

labs(title = "마력(hp)과 연비(mpg)의 관계 (실린더 개수별 색상)",

x = "마력 (hp)",

y = "연비 (mpg)",

color = "실린더 개수(cyl)") +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5),

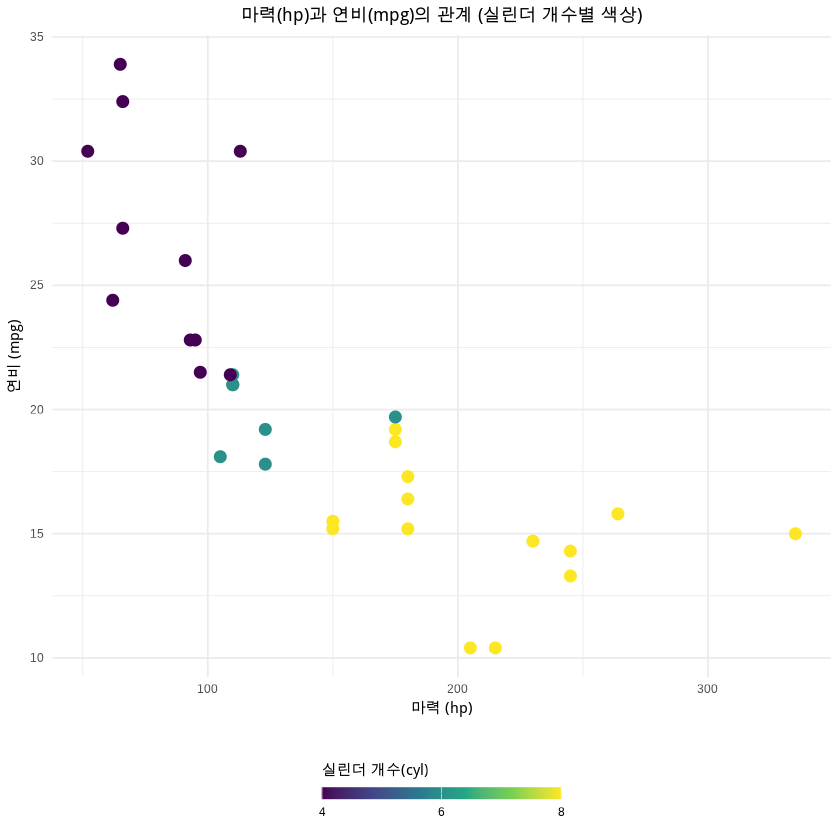
legend.position = "bottom", # 범례 위치 하단으로 조정 (수평 배치에 적합)

legend.box.margin = margin(t = 10)) # 그래프와 범례 사이 여백 추가

# 최종 그래프 출력 (R 환경에서 확인 및 캡처)

print(p\_mtcars\_final)

결과:



해석:

생성된 그래프(p\_mtcars\_final)는 1번 그래프의 요구사항을 만족하면서 추가적인 시각적 조정을 포함합니다. 점의 색상은 실린더 개수(cyl)를 나타내며, 색상 표현에는 'viridis' 컬러맵이 사용되어 색각 이상이 있는 사용자도 비교적 쉽게 구분할 수 있습니다. 또한, 색상과 실린더 개수를 연결하는 범례가 기본 세로 막대 형태가 아닌, 그래프 하단에 수평 방향의 컬러바로 표시됩니다. 이 컬러바에는 실제 실린더 값인 4, 6, 8이 명확히 표시되어 색상만 보고도 해당 자동차의 실린더 개수를 쉽게 파악할 수 있도록 돕습니다. 전반적인 레이아웃은 제목과 축 레이블이 명확하며, 수평 컬러바는 그래프 아래쪽에 위치하여 시각적 균형을 이룹니다.

3. 완성된 그래프를 통해 실린더 개수, 마력, 연비 간의 관계에 대해 자신의 견해를 서술하시오.

완성된 산점도 그래프를 통해 자동차의 실린더 개수(cyl), 마력(hp), 그리고 연비(mpg) 사이에는 다음과 같은 뚜렷한 관계가 있음을 알 수 있습니다.

마력(hp)과 연비(mpg)는 전반적으로 강한 음의 상관관계를 보입니다. 즉, 마력이 높은 자동차일수록 연비는 낮아지는 경향이 매우 뚜렷합니다. 그래프에서 점들이 전체적으로 우하향하는 패턴을 그리는 것을 통해 이를 확인할 수 있습니다.

실린더 개수(cyl)는 마력 및 연비와 밀접한 관련이 있습니다. 그래프에서 색상으로 구분된 실린더 그룹을 살펴보면 명확한 경향을 알 수 있습니다.

- 4기통(cyl=4) 자동차: 대부분 낮은 마력(약 50~120 hp 범위)과 높은 연비(약 20~35 mpg 범위) 영역에 집중되어 있습니다.

- 6기통(cyl=6) 자동차: 중간 정도의 마력(약 100~180 hp 범위)과 중간 정도의 연비(약 17~22 mpg 범위) 영역에 주로 분포합니다.

- 8기통(cyl=8) 자동차: 높은 마력(약 150~350 hp 범위)과 낮은 연비(약 10~20 mpg 범위) 영역에 주로 분포합니다.

이를 종합하면 실린더 개수가 많을수록 대체로 마력이 높고 연비는 낮아지는 경향이 있음을 명확히 알 수 있습니다. 실린더가 많다는 것은 일반적으로 엔진의 배기량이 크다는 것을 의미하며, 이는 더 높은 출력을 낼 수 있지만 동시에 연료 소모도 많아지기 때문으로 해석할 수 있습니다. 그래프 상에서 4, 6, 8기통 그룹이 비교적 명확하게 군집을 이루며 우하향 방향으로 순차적으로 이동하는 모습이 이러한 관계를 시각적으로 보여줍니다. 따라서 실린더 개수는 자동차의 성능(마력)과 효율성(연비)을 결정하는 중요한 요인 중 하나이며, 이 세 변수는 서로 밀접하게 연관되어 있다고 결론 내릴 수 있습니다.

**문제3.** [6강. 색(color)]

R의 내장 데이터셋 중 하나인 'iris'를 활용하여 꽃받침(Sepal)의 길이와 너비의 관계를 3품종(Species)별로 시각화하려고 한다. (40점)

(R에서 iris를 콘솔에 입력하거나 스크립트에 적어서 실행시키면 불러올 수 있음)

선행 코드:

# iris 데이터셋 불러오기

data(iris)

1. 'Sepal.Length'를 x축으로, 'Sepal.Width'를 y축으로 하는 산점도를 그리시오.

코드:

p\_iris\_base <- ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width)) +

geom\_point(size = 3) + # 점 크기 조절

labs(title = "붓꽃(iris) 꽃받침 길이와 너비의 관계",

x = "꽃받침 길이 (Sepal.Length)",

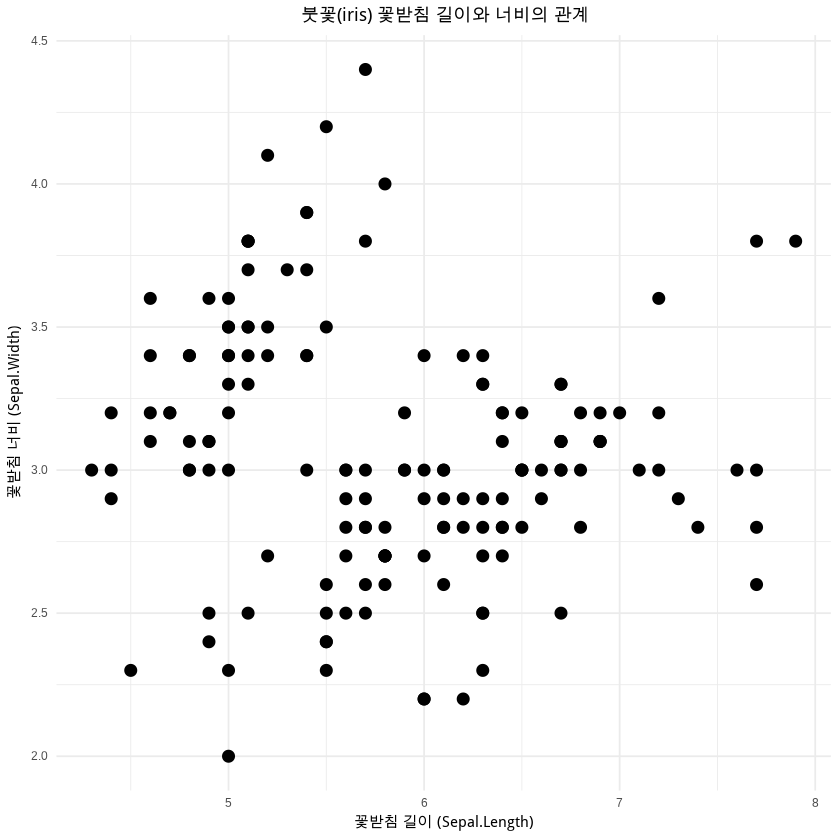
y = "꽃받침 너비 (Sepal.Width)") +

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5)) # 제목 가운데 정렬

print(p\_iris\_base)

결과:



해석:

각 점은 개별 붓꽃 샘플 하나를 나타냅니다. 이 그래프만으로는 붓꽃 품종(Species) 간의 차이를 구분할 수 없으며, 전체 붓꽃 데이터에 대해 꽃받침 길이와 너비 사이에 어떤 관계가 있는지 대략적으로 보여줍니다. 전반적으로 명확한 선형 관계가 보이지는 않지만, 데이터가 특정 영역에 군집하는 경향을 관찰할 수 있습니다.

2. 품종(Species)에 따라 점의 색상을 다르게 표현하기 위해 scale\_color\_brewer() 함수를 사용하여 'Set1' 색상 팔레트를 적용하시오.

코드:

p\_iris\_final <- ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +

geom\_point(size = 3) +

# scale\_color\_brewer()를 사용하여 'Set1' 팔레트 적용

scale\_color\_brewer(palette = "Set1") +

labs(title = "품종별 붓꽃(iris) 꽃받침 길이와 너비의 관계",

x = "꽃받침 길이 (Sepal.Length)",

y = "꽃받침 너비 (Sepal.Width)",

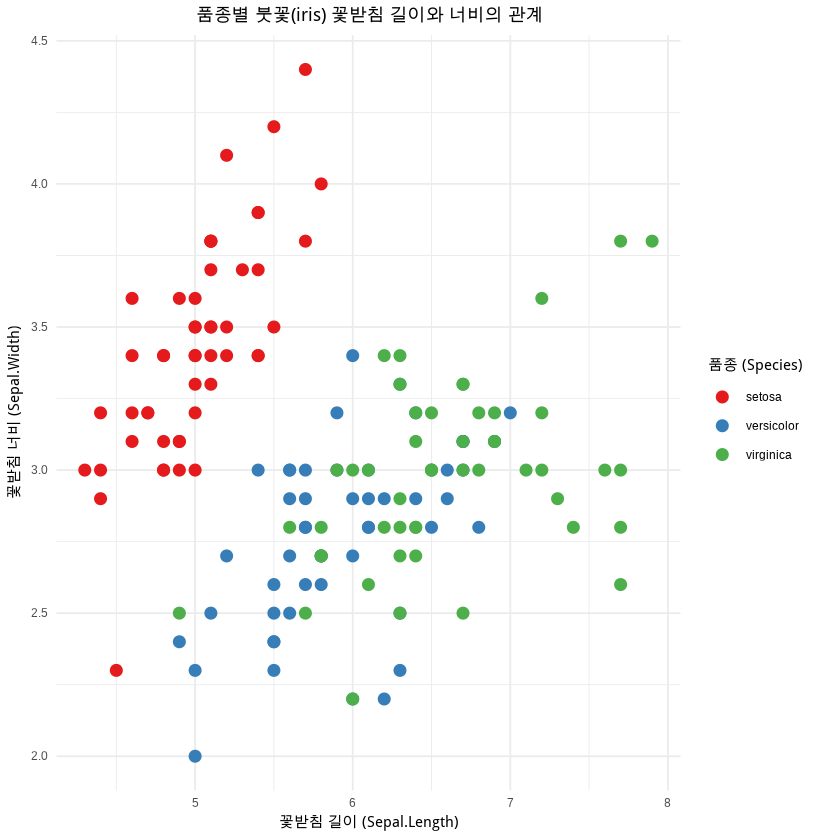
color = "품종 (Species)") + # 범례 제목

theme\_minimal() +

theme(plot.title = element\_text(hjust = 0.5)) # 제목 가운데 정렬

print(p\_iris\_final)

결과:



해석:

3. 생성된 그래프에 대해 품종별 꽃받침 길이와 너비의 관계를 설명하시오.

생성된 그래프를 통해 세 붓꽃 품종이 꽃받침 길이와 너비 측면에서 상당한 차이가 있음을 알 수 있습니다.

- Setosa 품종: 이 품종은 다른 두 품종과 명확히 구별되며, 꽃받침 길이는 가장 짧고 너비는 가장 넓은 특징을 보입니다. Setosa 그룹 내에서는 길이와 너비 사이에 뚜렷한 상관관계는 잘 보이지 않으며, 다른 두 품종과 명확히 구분되는 독립적인 군집을 형성하고 있습니다.

- Versicolor 품종: 이 품종은 꽃받침 길이와 너비 모두 중간 정도의 값을 가집니다. Versicolor 품종 내에서는 꽃받침 길이가 길어질수록 너비도 함께 넓어지는 비교적 뚜렷한 양의 상관관계가 나타납니다. Setosa와는 명확히 구분되지만, 일부 영역에서는 Virginica 품종과 분포가 다소 겹치는 것을 볼 수 있습니다.

- Virginica 품종: 이 품종은 꽃받침 길이가 가장 긴 특징을 가지며, 너비는 Versicolor와 비슷하거나 약간 더 넓은 범위에 분포합니다. 이 품종 내에서도 꽃받침 길이가 길수록 너비가 함께 증가하는 경향이 관찰됩니다. Versicolor 품종과 일부 겹치는 부분이 존재하지만, 전반적으로 더 긴 꽃받침 3길이를 갖는 영역에서 발견됩니다.

Setosa는 짧고 넓은 꽃받침, Virginica는 긴 꽃받침, 그리고 Versicolor는 그 중간적인 특성을 가집니다. 이를 통해 세 품종은 꽃받침의 길이와 너비 특성만으로도 상당히 잘 구분됨을 알 수 있습니다. 특히 Versicolor와 Virginica 품종은 꽃받침 길이가 길수록 너비도 함께 넓어지는 유사한 양의 상관관계를 보입니다.

**<과제물 작성 관련 유의사항>**

1. 모든 문항에 대한 답안을 하나의 파일로 작성하며, 이미지는 적정 크기로 하여 제출 문서 파일 내에 포함한다.

2. 모든 문항은 과제물 파일에 R 프로그램 코드와 실행한 결과물(캡처 이미지)을 같이 포함하여 작성한다.