Carseats Report - Advertising vs Sales

Learning Spoons R

2018-10-20

## I. Carseat

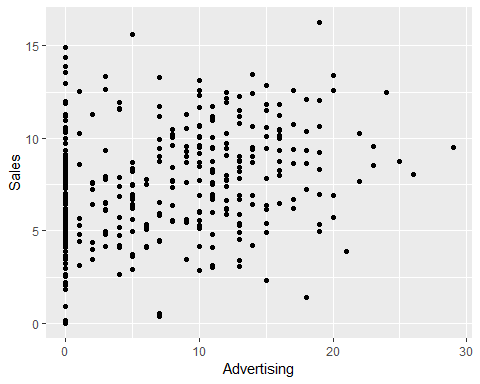
library(ISLR)  
library(dplyr)  
library(ggplot2)  
str(Carseats)

## 'data.frame': 400 obs. of 11 variables:  
## $ Sales : num 9.5 11.22 10.06 7.4 4.15 ...  
## $ CompPrice : num 138 111 113 117 141 124 115 136 132 132 ...  
## $ Income : num 73 48 35 100 64 113 105 81 110 113 ...  
## $ Advertising: num 11 16 10 4 3 13 0 15 0 0 ...  
## $ Population : num 276 260 269 466 340 501 45 425 108 131 ...  
## $ Price : num 120 83 80 97 128 72 108 120 124 124 ...  
## $ ShelveLoc : Factor w/ 3 levels "Bad","Good","Medium": 1 2 3 3 1 1 3 2 3 3 ...  
## $ Age : num 42 65 59 55 38 78 71 67 76 76 ...  
## $ Education : num 17 10 12 14 13 16 15 10 10 17 ...  
## $ Urban : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 2 2 1 1 ...  
## $ US : Factor w/ 2 levels "No","Yes": 2 2 2 2 1 2 1 2 1 2 ...

이 보고서는 Advertising이 Sales에 미치는 영향을 분석합니다.

## II. Basic: Advertising vs Sales

xyPlot <-   
 ggplot(data = Carseats, aes(x = Advertising, y = Sales)) +   
 geom\_point()  
print(xyPlot)



## III. Design

* 후보 추가 변수들은 다음과 같습니다. setdiff(colnames(Carseats), c("Sales", "Adverting")): CompPrice, Income, Advertising, Population, Price, ShelveLoc, Age, Education, Urban, US.
* 각각의 데이터는 각각의 지역 (Spatial Data)를 의미합니다. 그리고 Advertising와 Sales를 제외한 다른 변수들은 1) 해당 지역의 특성 혹은 2) 해당 지역에서의 business practice의 특성을 반영합니다. (factor 변수는 bold로 표기하였습니다.)

1. 해당 지역의 특성: Income, Population, Age, Education, **Urban**, **US**
2. 해당 지역의 biz의 특성: CompPrice, Price, **ShelveLoc**

* 해당 지역의 특성에 따라서 광고 효과를 살펴봅니다.

1. Factor 변수에 따라서 광고 효과가 달라지는지 살펴봅니다.
2. Factor 변수가 아닌 경우에는 4분위 값을 기준으로 Factor 변수로 변환하여 살펴봅니다.

## IV. Report

### 1. Factor 변수 분석