




# Day24; 20221007

📅 날짜	
📅 유형	
🏷 태그	

GitHub - u8yes/R

You can't perform that action at this time. You signed in with another tab or window. You signed out in another tab or window. Reload to refresh your session. Reload to refresh your session.

 <https://github.com/u8yes/R>

u8yes/R



 1 Contributor  
 0 Issues  
 1 Star  
 0 Forks



[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/6f92dd73-907c-4bfc-a3bd-d54119649ab5/01.\\_R\\_%EC%84%A4%EC%B9%98\\_%EB%B0%8F\\_%EA%B0%9C%EC%9A%94.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/6f92dd73-907c-4bfc-a3bd-d54119649ab5/01._R_%EC%84%A4%EC%B9%98_%EB%B0%8F_%EA%B0%9C%EC%9A%94.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/662f9add-c8b5-41a3-a159-a9c0c1fee022/02.\\_%EB%8D%B0%EC%9D%B4%ED%84%B0\\_%EC%9C%A0%ED%98%95%EA%B3%BC\\_%EA%B5%AC%EC%A1%B0.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/662f9add-c8b5-41a3-a159-a9c0c1fee022/02._%EB%8D%B0%EC%9D%B4%ED%84%B0_%EC%9C%A0%ED%98%95%EA%B3%BC_%EA%B5%AC%EC%A1%B0.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/11bf7951-f61b-4d28-82fa-6c71c9ed0b18/03.\\_%EB%8D%B0%EC%9D%B4%ED%84%B0\\_%EC%9E%85%EC%B6%9C%EB%A0%A5.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/11bf7951-f61b-4d28-82fa-6c71c9ed0b18/03._%EB%8D%B0%EC%9D%B4%ED%84%B0_%EC%9E%85%EC%B6%9C%EB%A0%A5.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/cec92af5-b217-49cd-bd75-ccb9c94f81b7b/04.\\_%EC%A0%9C%EC%96%B4%EB%AC%B8%EA%B3%BC\\_%ED%95%A8%EC%88%98.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/cec92af5-b217-49cd-bd75-ccb9c94f81b7b/04._%EC%A0%9C%EC%96%B4%EB%AC%B8%EA%B3%BC_%ED%95%A8%EC%88%98.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/063c6423-cf04-427c-b5d9-63992b0aaee2/05.\\_%EB%8D%B0%EC%9D%B4%ED%84%B0\\_%EC%8B%9C%EA%B0%81%ED%99%94.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/063c6423-cf04-427c-b5d9-63992b0aaee2/05._%EB%8D%B0%EC%9D%B4%ED%84%B0_%EC%8B%9C%EA%B0%81%ED%99%94.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/6ec901ee-1dde-446c-817f-71f2fd89f34b/R\\_%EC%88%98%EC%97%85\\_%EA%B4%80%EB%A0%A8\\_%EB%8D%B0%EC%9D%B4%ED%84%B0\\_%EC%85%8B\\_%EC%84%A4%EB%AA%85.pdf](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/6ec901ee-1dde-446c-817f-71f2fd89f34b/R_%EC%88%98%EC%97%85_%EA%B4%80%EB%A0%A8_%EB%8D%B0%EC%9D%B4%ED%84%B0_%EC%85%8B_%EC%84%A4%EB%AA%85.pdf)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/0a9a8e84-98d9-44e0-ba25-60b1175ef455/chap01\\_Basic.r](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/0a9a8e84-98d9-44e0-ba25-60b1175ef455/chap01_Basic.r)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/afe7eda7-43cd-4f98-a575-5cefc9a0698c/chap02\\_DataStructure.r](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/afe7eda7-43cd-4f98-a575-5cefc9a0698c/chap02_DataStructure.r)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/9cc3d5b9-5102-4962-bd45-64f918f81c1a/chap03\\_DataIO.r](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/9cc3d5b9-5102-4962-bd45-64f918f81c1a/chap03_DataIO.r)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/aa206225-1806-4ba8-a9e0-b79a9373efc5/chap04\\_Control-Function.r](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/aa206225-1806-4ba8-a9e0-b79a9373efc5/chap04_Control-Function.r)

[https://s3-us-west-2.amazonaws.com/secure.notion-static.com/513327bd-b7d9-45ae-987b-bd9b6d458b30/chap05\\_DataVisualization.r](https://s3-us-west-2.amazonaws.com/secure.notion-static.com/513327bd-b7d9-45ae-987b-bd9b6d458b30/chap05_DataVisualization.r)

---

난수 - 임의로 뽑는, 규칙적이지 않은 숫자

---

종 모양은 평균에 가장 데이터가 몰려있다는 것. (좌우 대칭의 분포 모양) = 정규 분포

---

데이터가 얼마나 흩어져있는지 알 수가 없다.

내 값과 평균의 차 - 편차

---

대표값(평균)에 제곱을 씌워서 오차값을 극대화해서 평균을 계산 - 분산

---



평균을 0으로, 표준편차를 1로

종모양에서 가운데가 0, 양쪽으로 퍼지는 게 1.

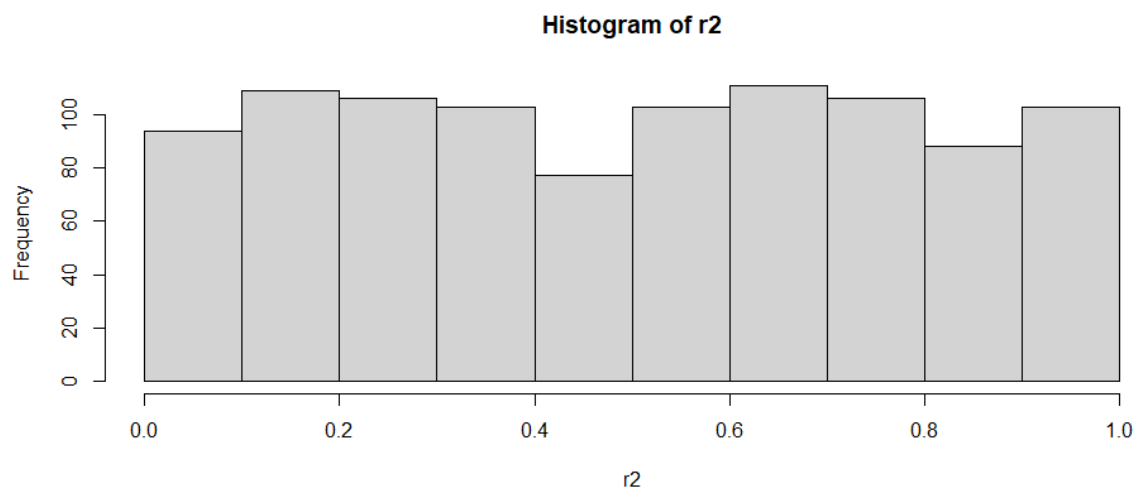
이러한 특징을 가진 것이 표준 정규분포.

---

```
# 단계2 : 균등분포(연속형)의 난수 생성
n <- 1000
r2 <- runif(n, min = 0, max = 1) # 0 < r2 < 1 # R uniform 균등한 분포
hist(r2)
```

**uniform** 미국식 [ˈjuːnɪfɔːrm]  영국식 [ˈjuːnɪfɔːm]  ★★

1. [명사] 제복, 군복, 교복, 유니폼
2. [형용사] 획일적인, 균일한, 한결같은



---

수집 → 저장 → (전)처리 → 분석

(전)처리 - 데이터 이해도 등

분석 - 알고리즘, 시각화 등

---

Client 가 데이터를 의뢰할 때는 중구난방으로 주기 때문에 전처리 과정을 통해 찾아내야 한다.

Hadoop 기술 - 하지만 너무 어려움(서버만 있으면 무료)

---

어떤 데이터를 만날지 모름. 평소 용어에 대한 데이터 이해도 과정이 많이 소요된다.

데이터를 가지고 바로 결과를 내려고 하면 정확도가 떨어질 수밖에 없다.

당연히 전처리 과정을 거쳐야 좋은 분석 시각화가 유의미한 결과를 도출할 수 있다.

분석가가 얼마나 이해하고 파악하는냐인데 실무자(담당자)와의 많은 소통을 해야 한다.

아주 짧은 시간 안에...

실무자는 잘 가르쳐주지 않는다.

데이터가 어디있는지도 안 가르쳐준다.

---

1차원 vector

```

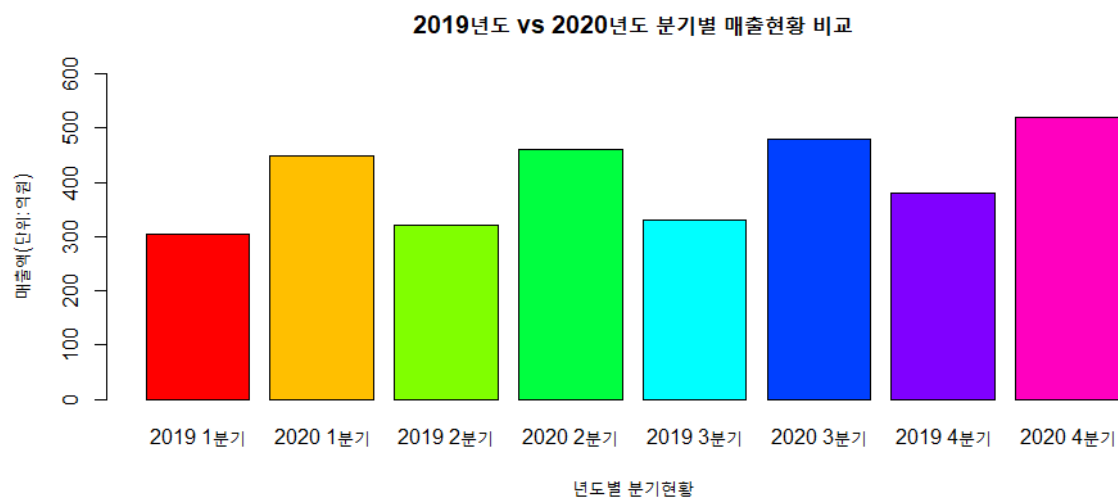
14
15 # 막대 차트 데이터 생성
16 chart_data <- c(305, 450, 320, 460, 330, 480, 380, 520) # 정수 데이터
17 names(chart_data) <- c("2019 1분기", "2020 1분기", "2019 2분기", "2020 2분기",
18 "2019 3분기", "2020 3분기", "2019 4분기", "2020 4분기")
19
20 str(chart_data)
21 chart_data
22
23 # 세로 막대 차트
24 help("barplot")
25 help(barplot)
26 barplot(chart_data, ylim = c(0, 600), col = rainbow(8), main = "2019년도 vs
27 2020년도 분기별 매출현황 비교")
28
29 barplot(chart_data, ylim = c(0, 600), col = rainbow(8),
30         xlab = "년도별 분기현황", ylab = "매출액(단위:만원)",
31         main = "2019년도 vs 2020년도 분기별 매출현황 비교")
32 # barplot(vector, vlimit(0부터 600까지만), col=rainbow(자동 8개의 색상으로),

```

```

> chart_data
2019 1분기 2020 1분기 2019 2분기 2020 2분기 2019 3분기 2020 3분기 2019 4분기
      305      450      320      460      330      480      380
2020 4분기
      520
> # 세로 막대 차트

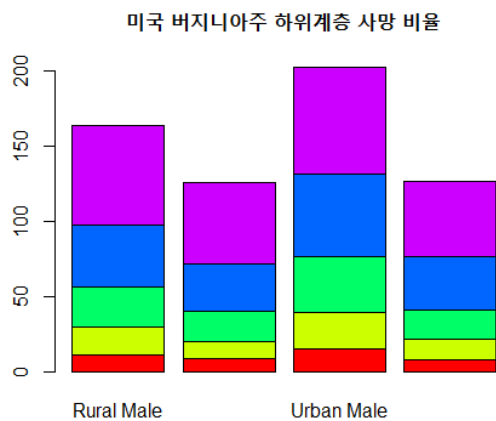
```



list 는 key, value라서 mode()를 하면 list라고 뜬.

## 2차원 Matrix

```
74  
75 # 누적 차트  
76 par(mfrow=c(1,2)) # 1행 2열 그래프 보기  
77  
78 barplot(VADeaths, col=rainbow(5),  
79         main = "미국 버지니아주 하위계층 사망 비율")  
80  
81 legend(3.8, 200,  
82        c("50-54", "55-59", "60-64", "65-69", "70-74"),  
83        cex = 0.8, fill = rainbow(5))  
84
```



시각화 CharacterEXpansion

barplot

```

par(mfrow=c(1,2)) # 1행 2열 그래프 보기

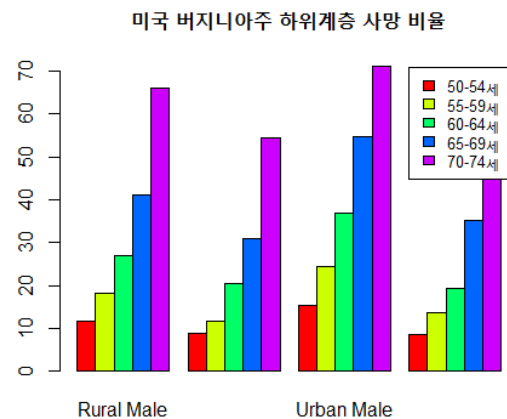
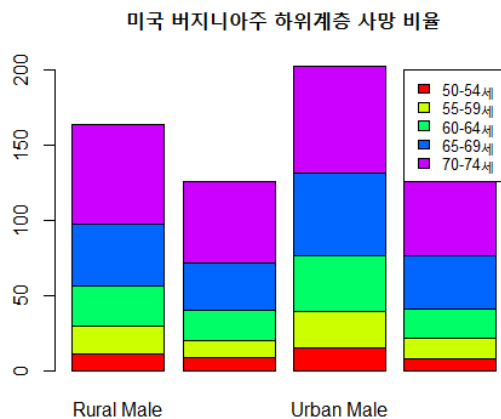
barplot(VADeaths, col=rainbow(5),
        main = "미국 버지니아주 하위계층 사망 비율")

legend(3.8, 200,
       c("50-54세", "55-59세", "60-64세", "65-69세", "70-74세"),
       cex = 0.8, fill = rainbow(5)) # 3.8 위치

# 개별 차트
barplot(VADeaths, col=rainbow(5),
        main = "미국 버지니아주 하위계층 사망 비율",
        beside=T) # 개별적으로 나누어서 시각화

legend(19, 71,
       c("50-54세", "55-59세", "60-64세", "65-69세", "70-74세"),
       cex = 0.8, fill = rainbow(5))

```



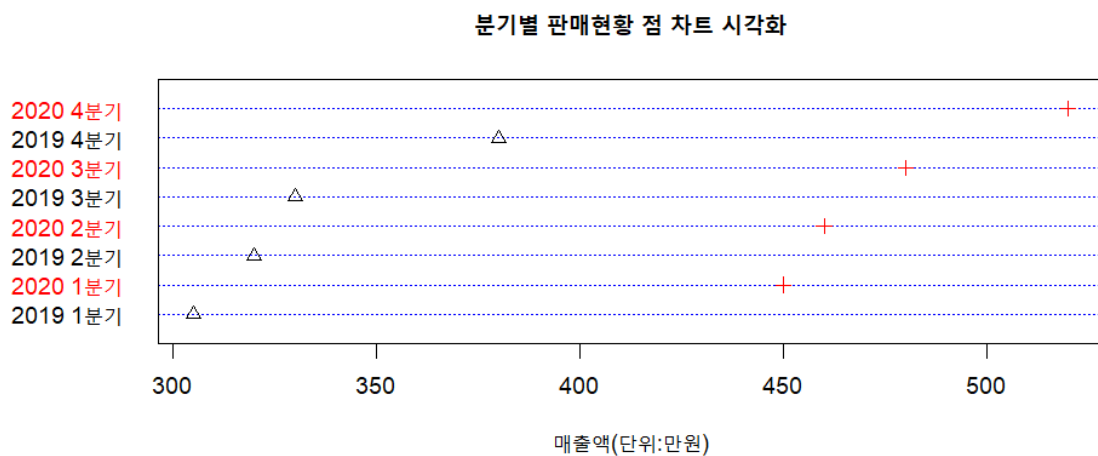
dotchart



```

105
106 dotchart(chart_data, color = c("black", "red"),
107           xlab = "매출액(단위:만원)",
108           cex=1.2,
109           main="분기별 판매현황 점 차트 시각화",
110           labels=names(chart_data),
111           lcolor="blue", pch=2:3)
112 # pch(plotting character): 원(1), 삼각형(2), +(3)
113 # cex(character expansion): 레이블과 점의 크기 확대 역할.
114

```

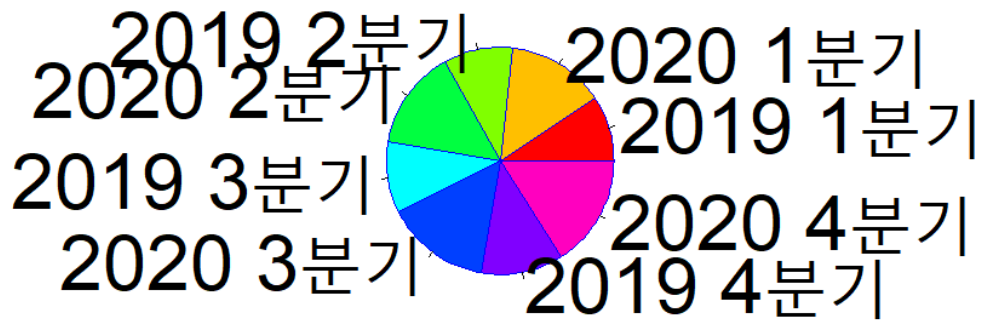


pie

```

16 help(pie)
17
18 pie(chart_data, labels = names(chart_data),
19      border = 'blue', col=rainbow(8), cex=4)
20 title("2019~2020년도 분기별 매출현황")
21

```



boxplot 상자수염그림

```

127 help(boxplot)
128 par(mfrow=c(1,2))
129 boxplot(VADeaths) # 상자그래프 시각화.
130 boxplot(VADeaths, range=0)
131 # range=0:최소값과 최대값을 점선으로 연결하는 역할.
132
133 abline(h = 37, lty=3, col="red") # 기준선 추가(lty=3 : 점선)
134
135 summary(VADeaths) # median이 2사분위
136

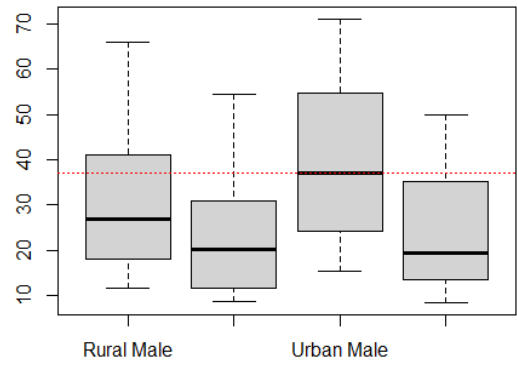
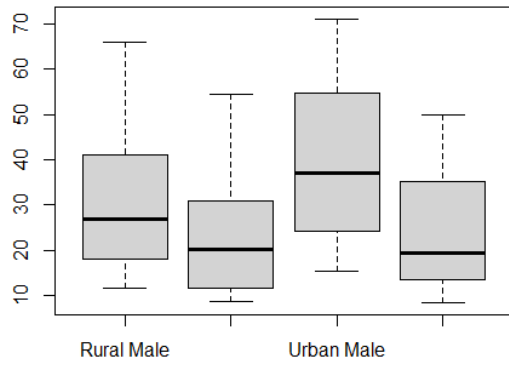
```

```

> par(mfrow=c(1,2))
> boxplot(VADeaths) # 상자그래프 시각화.
> boxplot(VADeaths, range=0)
> abline(h = 37, lty=3, col="red") # 기준선 추가(lty=3 : 점선)
> summary(VADeaths)

```

Rural Male	Rural Female	Urban Male	Urban Female
Min. :11.70	Min. : 8.70	Min. :15.40	Min. : 8.40
1st Qu.:18.10	1st Qu.:11.70	1st Qu.:24.30	1st Qu.:13.60
Median :26.90	Median :20.30	Median :37.00	Median :19.30
Mean :32.74	Mean :25.18	Mean :40.48	Mean :25.28
3rd Qu.:41.00	3rd Qu.:30.90	3rd Qu.:54.60	3rd Qu.:35.10
Max. :66.00	Max. :54.30	Max. :71.10	Max. :50.00



iris 붓꽃

