

The background of the slide features a variety of 3D statistical charts. On the left, there is a large 3D pie chart with several slices in different colors (yellow, green, blue, purple, orange). To the right of the pie chart and in the background, there are several 3D bar charts with bars of various heights and colors (orange, white, pink, purple, blue, green). In the foreground, there is a 3D line chart with a yellow line and a 3D stacked bar chart with blue and grey bars.

통계학 실습

6. R 튜토리얼

- R 언어는 통계 및 그래프 작업을 위한 프로그래밍 언어이다.
- 수많은 통계 관련 패키지가 이미 개발되어 있으며, 쉽게 익힐 수 있다.
- 의학, 수학 등 통계가 필요한 다양한 학문에서 이미 많이 쓰이고 있다.

- 아래의 링크에서 운영체제에 맞도록 다운로드 한다.
 - R 홈페이지 : <http://cran.nexr.com/>
 - 해당 사이트에서 튜토리얼이나 함수 설명, 추가 패키지 또한 볼 수 있다.
- Windows는 자체적으로 프로그래밍할 수 있는 툴을 포함하고 있다.
 - R 아이콘을 실행하면 프로그래밍이 가능하게 된다.
- MAC은 터미널을 이용하여 프로그래밍을 할 수 있다.
 - 터미널에서 'R' 을 입력하면 일련의 문구가 나오고 프로그래밍이 가능하게 된다.

```
> x = 7  
> x  
[1] 7  
> class(x)  
[1] "numeric"
```

- R에서는 실행문 뒤에 세미콜론(;)을 붙이지 않아도 된다.
 - 한 줄에 여러 실행문을 쓸 경우 필요하다.
- 변수명을 입력하면 바로 출력된다.
 - 출력에서 [1]은 행의 가장 앞에 있는 값의 index를 뜻한다. 한 줄에 10개씩 자료가 있을 경우, 각 줄에 [1], [11], [21], ...이 생기게 된다.
- 해당 변수의 타입이 궁금할 때는 class() 함수를 이용하면 된다.

Character 자료형

```
> x = as.character(3.1415)
```

```
> x
```

```
[1] "3.14"
```

```
> class(x)
```

```
[1] "character"
```

- 문자열 출력에는 항상 양 끝에 따옴표가 붙는다.
- 다른 프로그래밍 언어의 'char' 는 1바이트의 한 글자이지만, R에서는 문자열을 뜻한다.
- 다른 프로그래밍 언어와 같이 문자열을 다루는 여러 가지 함수가 존재한다.

Character 자료형의 함수

```
> fname = "James"
> lname = "Smith"
> paste(fname, lname)
[1] "James Smith"
> sprintf("My name is %s %s", fname, lname)
[1] "My name is James Smith"
> substr(paste(fname, lname), start=3, stop=5)
[1] "mes"
> sub("s", "z", paste(fname, lname))
[1] "Jamez Smith"
> |
```

- paste
 - 문자열을 순서대로 이어 붙인다.
- sprintf
 - 포맷(형식)에 맞추어 출력해준다.
 - %d : 정수, %f : 소수, %s : 문자열
- substr
 - 문자열에서 일정 부분 추출한다.
- sub
 - 해당하는 문자열을 주어진 문자열로 대체한다. 다른 언어에서 replace 역할을 한다.

외부에서 데이터 불러오기

```
> data <- read.csv("C:\\Users\\Teasung\\Desktop\\test.csv")
> data
  x y
1 1 1
2 2 2
3 3 3
4 4 5
5 5 8
> |
```

- CSV는 가장 흔히 접하게 되는 데이터셋이다.
 - R에서 엑셀, TXT, SPSS, SAS 또한 외부 데이터로 불러올 수 있다.
- read.csv 함수를 이용하여 외부에서 CSV 데이터를 불러온다.
 - ‘데이터셋변수명\$컬럼명’ 을 이용하여 각 열에 접근할 수 있다.

외부로 데이터 내보내기

```
> data <- 5:30
> data
[1] 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
> write.csv(data, "C:\\Users\\Teasung\\Desktop\\data.csv")
> |
```

- write.csv 함수를 이용하여 만들어둔 데이터를 파일로 저장할 수 있다.
- 기본적으로 프로그램 종료 시에 데이터들을 저장해둘 것인지 묻는다.
 - 프로그램을 다시 실행하면, 이전에 저장하고 이용하였던 데이터들을 사용 가능하다.

R 패키지 사용하기

불러오기만 하면 되는 패키지

```
> library(MASS)
```

새로 설치하여야 하는 패키지

```
> install.packages( "ggplot2" )
```

```
> library(ggplot2)
```

■ 프로그램 설치 시 다운로드 되는 패키지 외에 추가적인 패키지를 쓸 수 있다.

- 미리 다운로드 된 패키지 외에 사용 방법에 따라 두 종류가 존재한다.

■ 불러오기만 하면 되는 패키지

- 설치 때 다운로드 되어 추가 설치가 필요하지 않다.
- library(package)로 사용 가능하다.

■ 새로 설치하여야 하는 패키지

- install.packages(package) 함수를 이용하여 개별로 설치하여야 한다.
- library(package)로 사용 가능하다.

```
> library(MASS)
> str(Cars93)
```

- ‘MASS’ 패키지는 이미 설치되어 있기 때문에 불러오기만 하면 된다.
- MASS 패키지에 저장되어 있는 Cars93 데이터 프레임을 확인해보자.
- str() 함수를 이용하여 해당 패키지의 내용을 정리된 상태로 볼 수 있다.
 - 해당 변수의 자료형
 - 해당 변수의 의미

문제

```
> library(MASS)
> str(Cars93)
### Print..
> Cars93$Weight
### Print..
```

- MASS 패키지의 Cars93 패키지를 이용하여 자동차들의 무게를 확인하자.
- 주어진 DATA SET의 Variable에 접근하는 방법은 'DataSet\$Variable' 로 접근할 수 있다.
- 해당 접근된 데이터의 내용을 실제 파일로 내보내어 다룰 수 있도록 하자.
 - 앞서 배운 데이터 내보내기를 이용하여 바탕화면에 .csv 파일로 저장 후 검사