

A wide-angle photograph of the Matterhorn mountain in the Swiss Alps. The mountain is a prominent, sharp peak covered in snow and ice. In the foreground, there's a valley with a winding road and some snow-covered slopes. The sky is clear with a few wispy clouds.

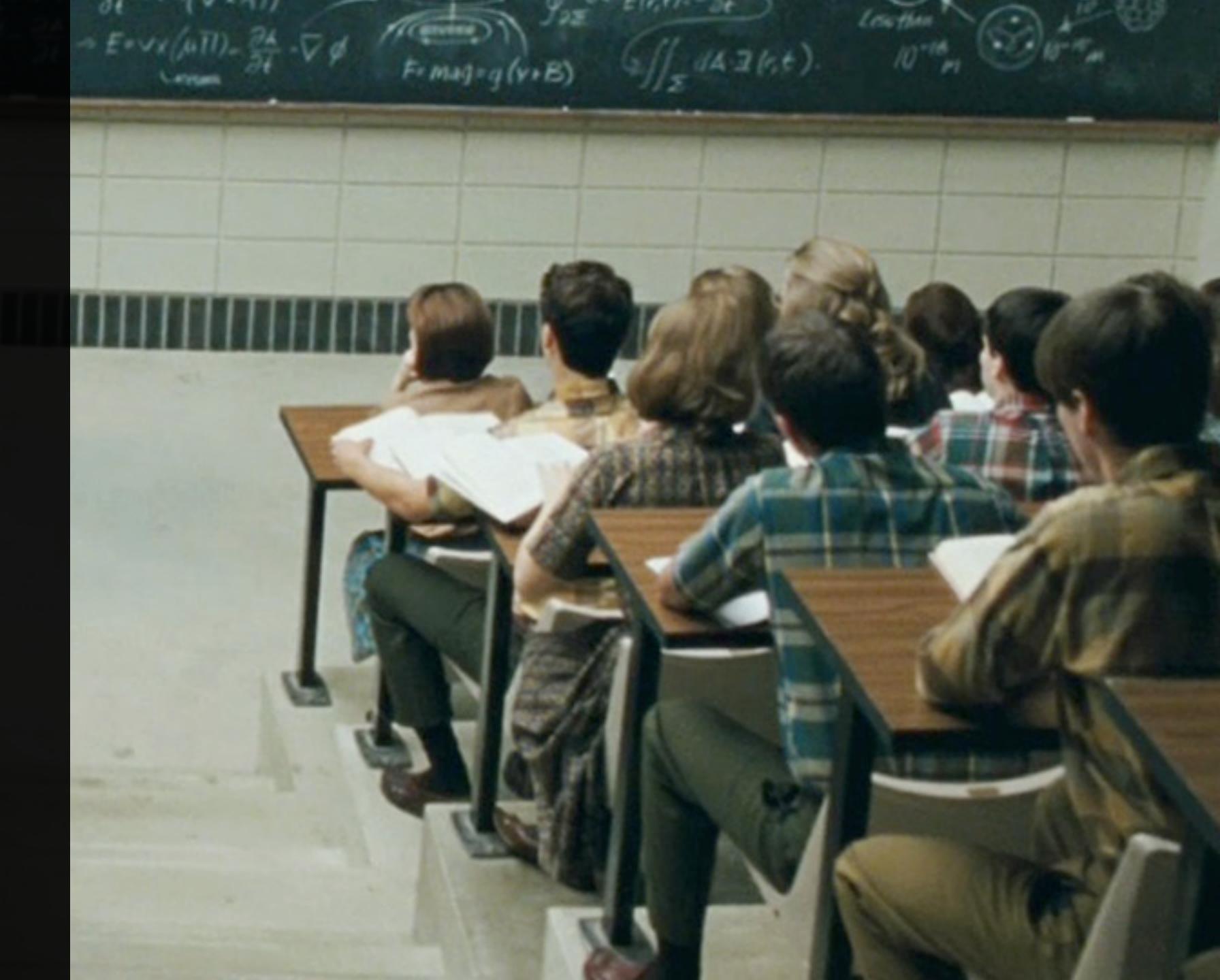
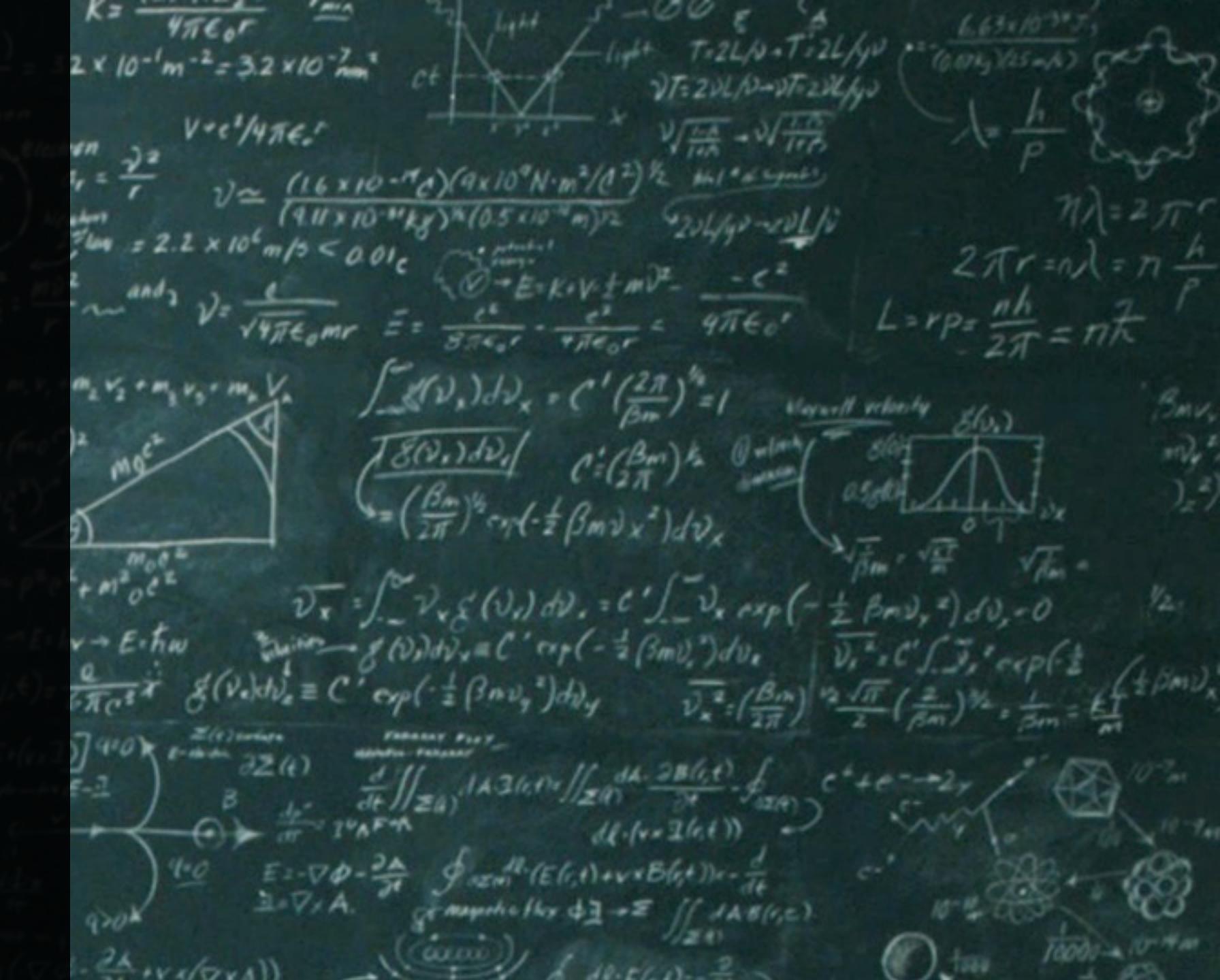
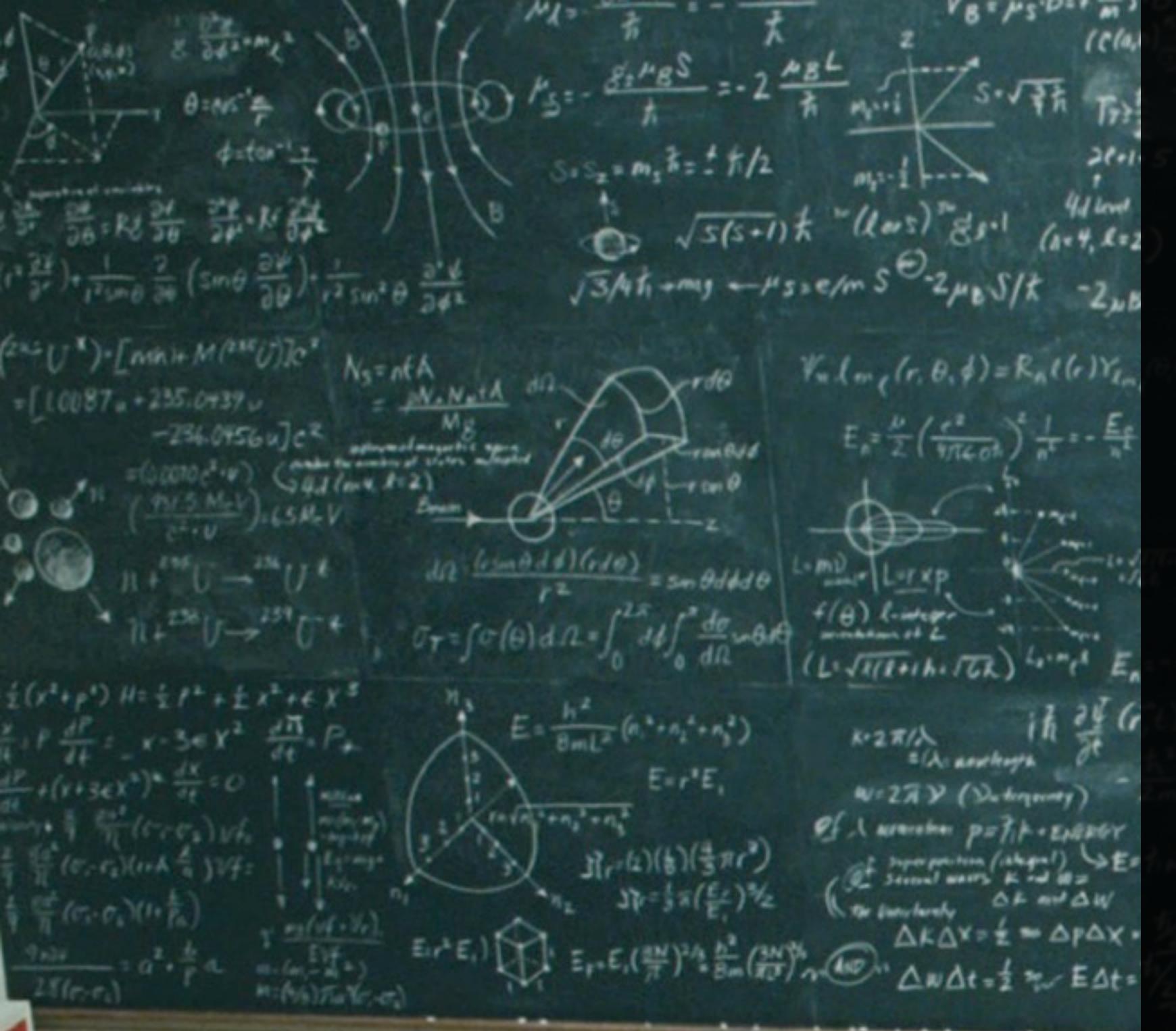
Big Leader
2016 Summer



WELCOME TO
AN INTRODUCTION TO MACHINE LEARNING WITH R



ABOUT THIS COURSE



HAN JAEYOON

otzslayer@gmail.com

otzslayer.github.io

facebook.com/otzslayer

B.S. Mathematics, Kyung Hee University, 2010 - Feb 2016

Master's Course, Kyung Hee University, Mar 2016 - Present

Data Fanatic

R User

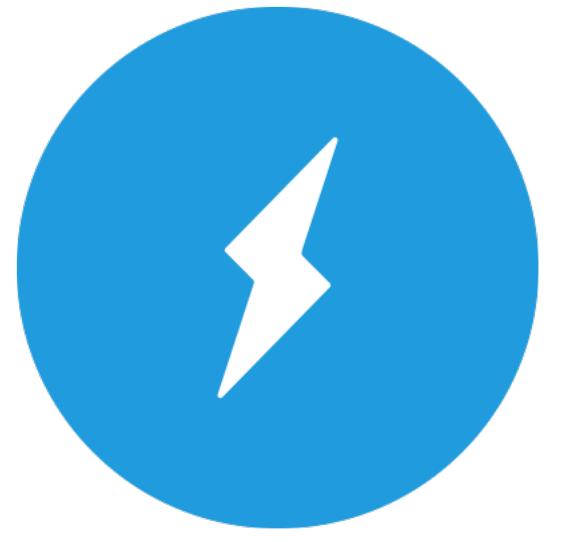
Kaggle (Top 10% @ Kobe Shot Selection Competition)

Research Interests

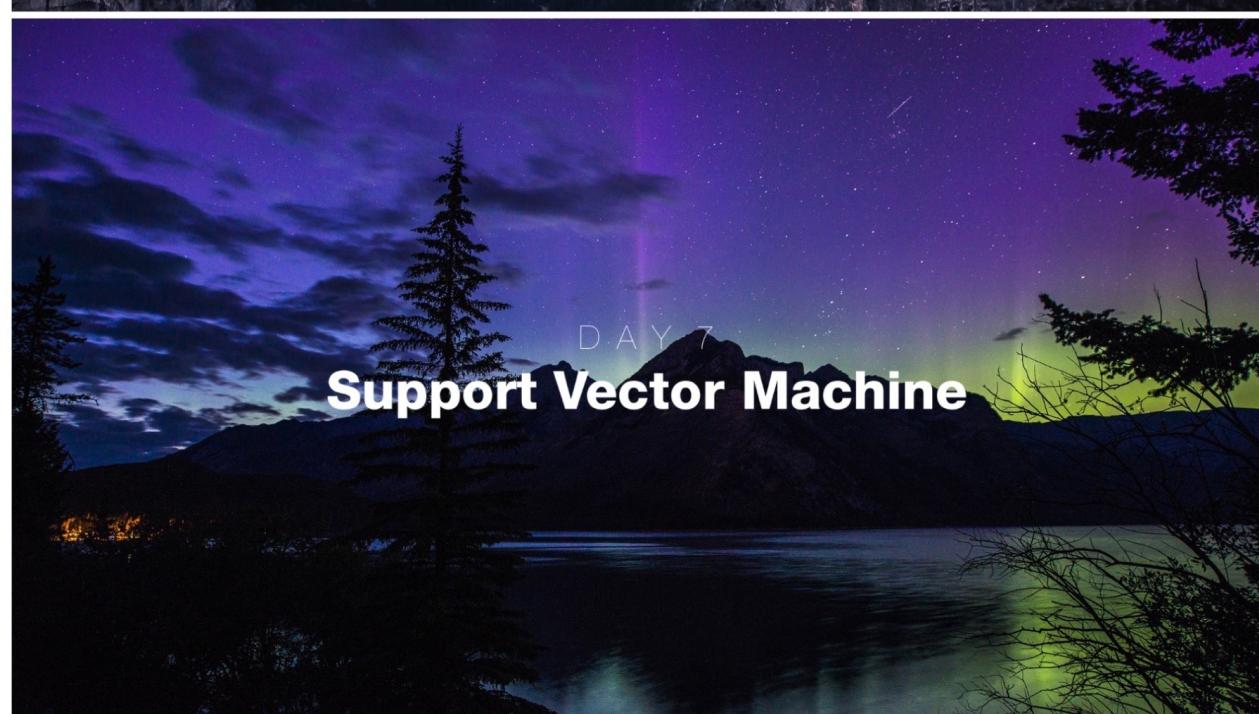
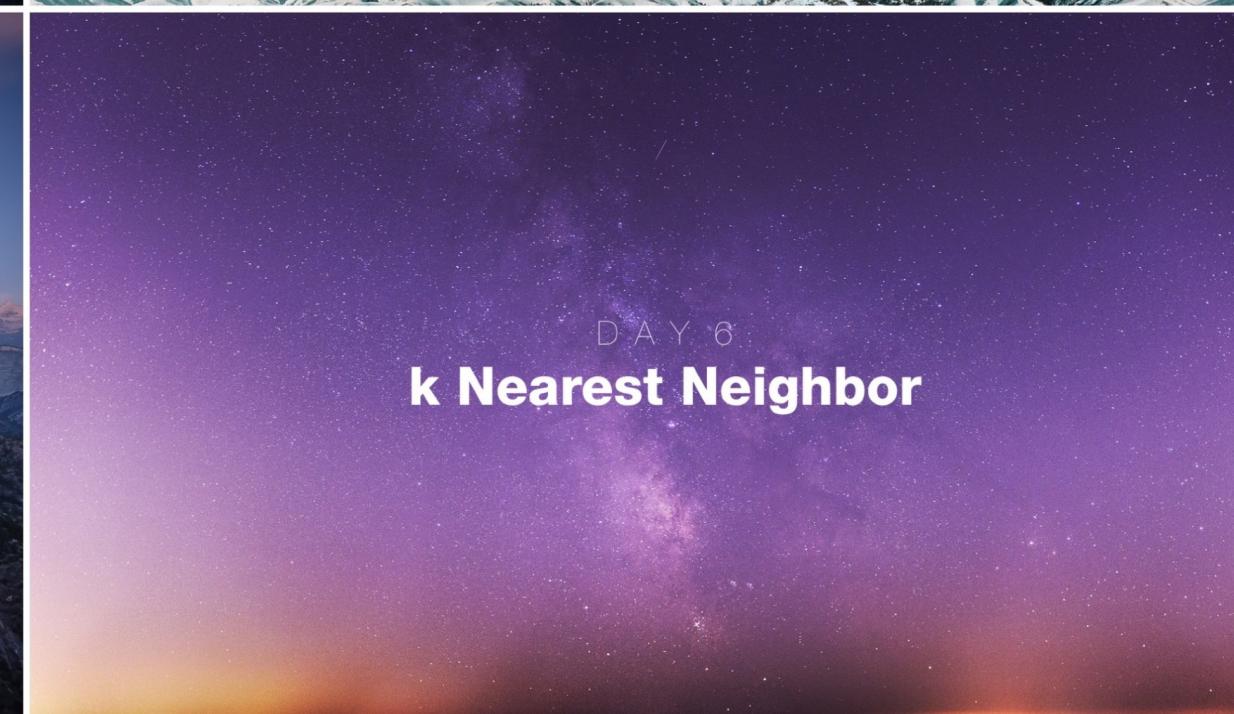
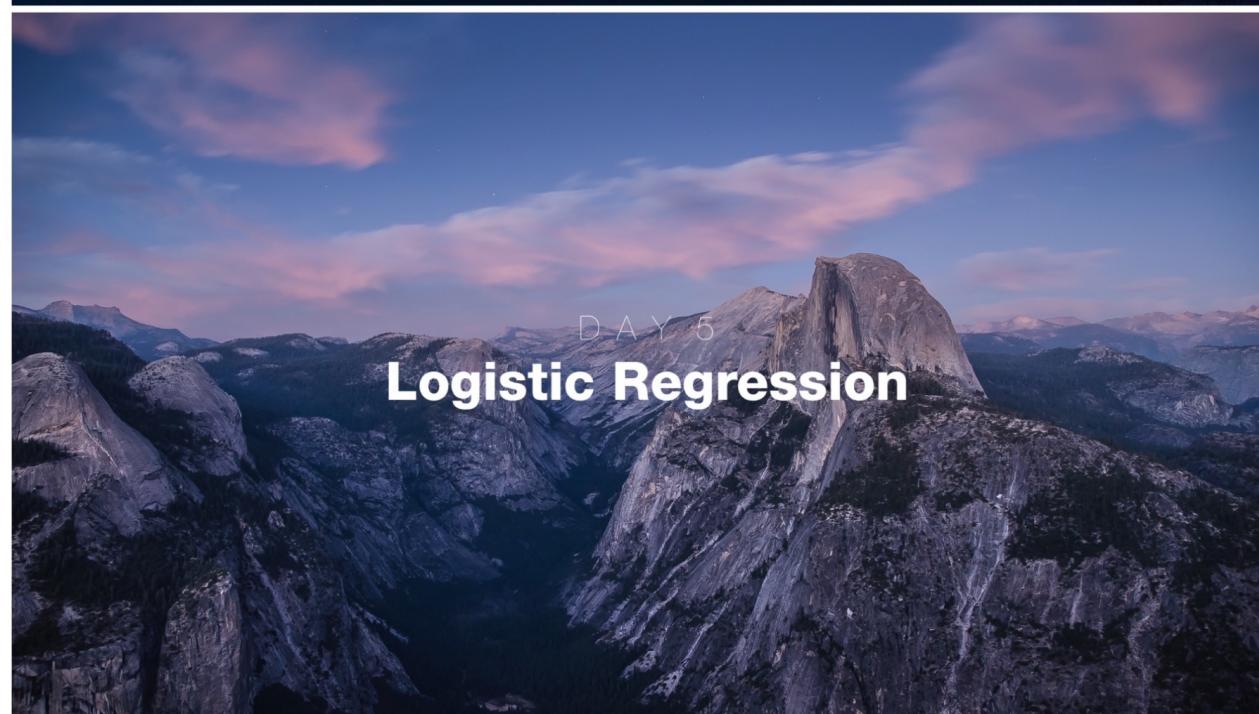
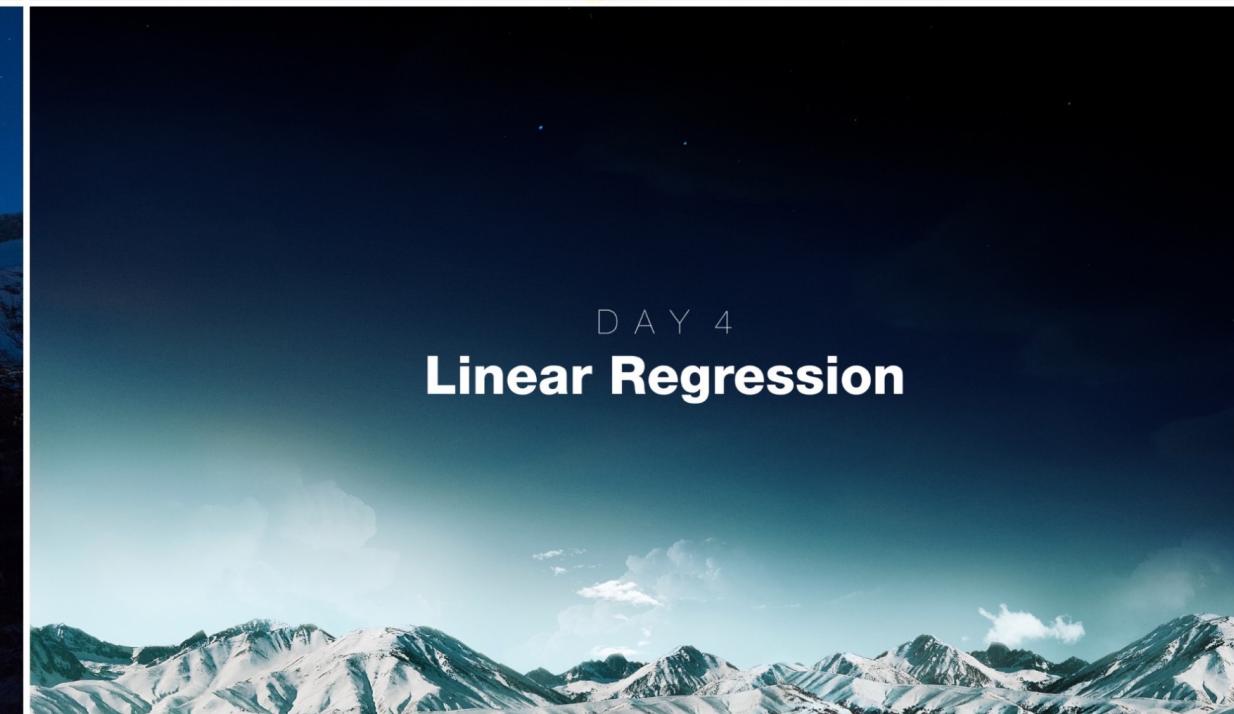
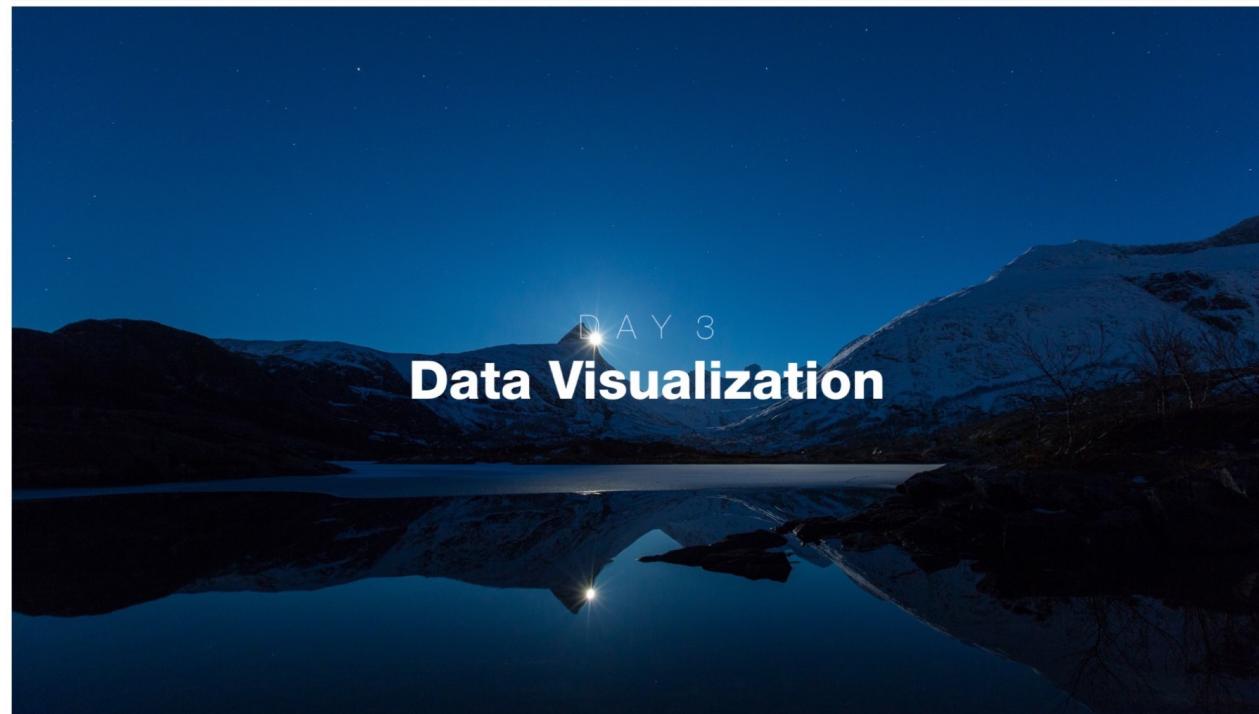
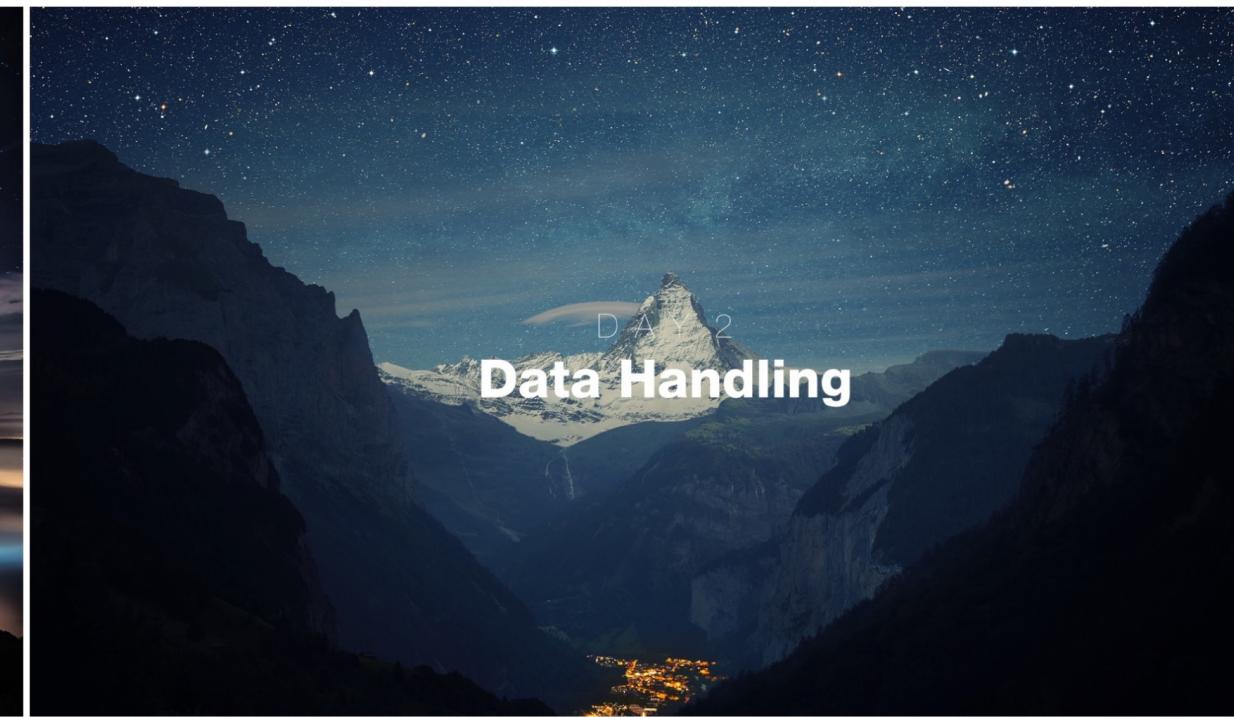
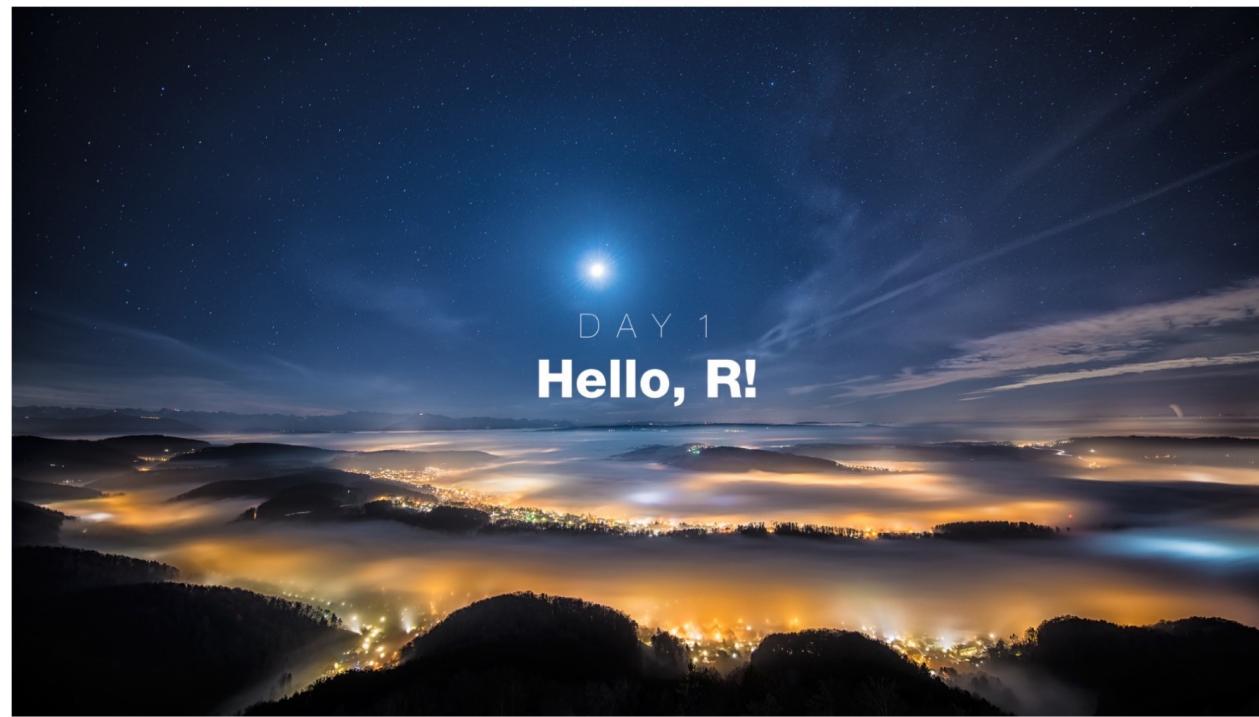
: Statistical Learning, Machine Learning, Deep Learning



PURPOSE OF THIS COURSE



FEATURES OF THIS COURSE



*BEAUTIFUL
KEYNOTE :P*

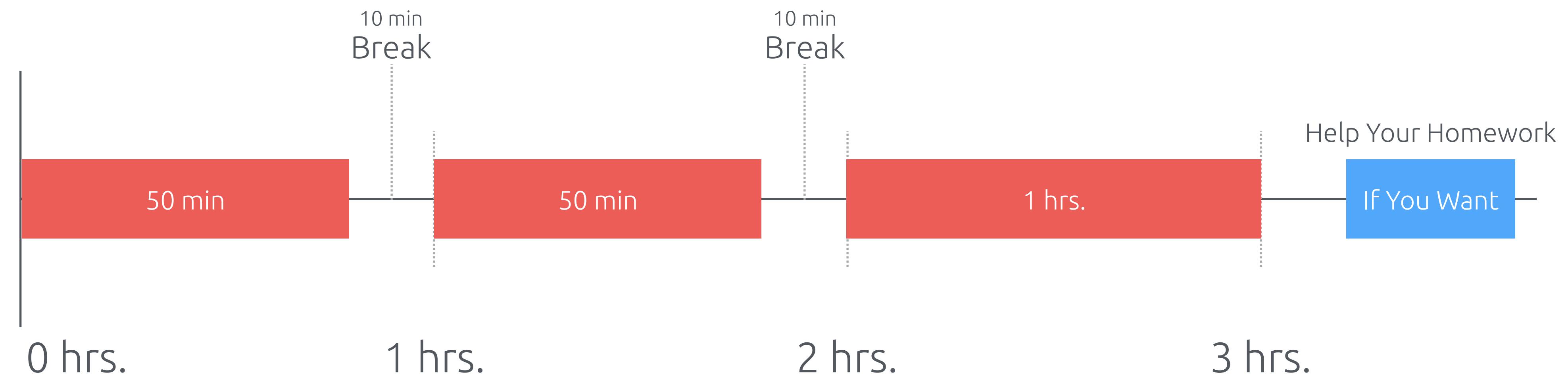
Less Math, But More Concept

Be Interactive

Be self-taught with Challenging Homework

CURRICULUM OF THIS COURSE

	Day	Date	Content
Week 1	Day 1	2016. 7. 19	Introduction
	Day 2	2016. 7. 22	Data Handling
Week 2	Day 3	2016. 7. 26	Data Visualization Intro for Machine Learning
	Day 4	2016. 7. 29	Linear Regression
Week 3	Day 5	2016. 8. 2	Logistic Regression
	Day 6	2016. 8. 5	K Nearest Neighbor Clustering
Week 4	Day 7	2016. 8. 9	Tree-Based Methods
	Day 8	2016. 8. 12	Support Vector Machine
Week 5	Day 9	2016. 8. 16	Kaggle Day 1
	Day 10	2016. 8. 19	Kaggle Day 2



A scenic landscape featuring a running track in the foreground, leading towards a vast green field and distant hills under a dramatic, cloudy sky.

**START
NOW**



AN
INTRODUCTION
TO
MACHINE
LEARNING
WITH R

DAY 1

A wide-angle night photograph of a valley. The foreground is dark with silhouettes of trees and hills. Below the hills, a layer of fog or low clouds glows with a warm orange and yellow light from the numerous small lights of a town or city nestled in the valley. In the background, a range of mountains is visible under a dark blue sky. A single, very bright star or the moon is positioned in the upper center of the frame, casting a soft glow. The overall atmosphere is serene and majestic.

DAY 1
Hello, R!



[데이터란?]

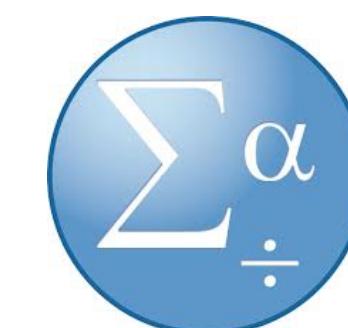
WHAT IS DATA?



[엑셀]
EXCEL



[SAS]
SAS



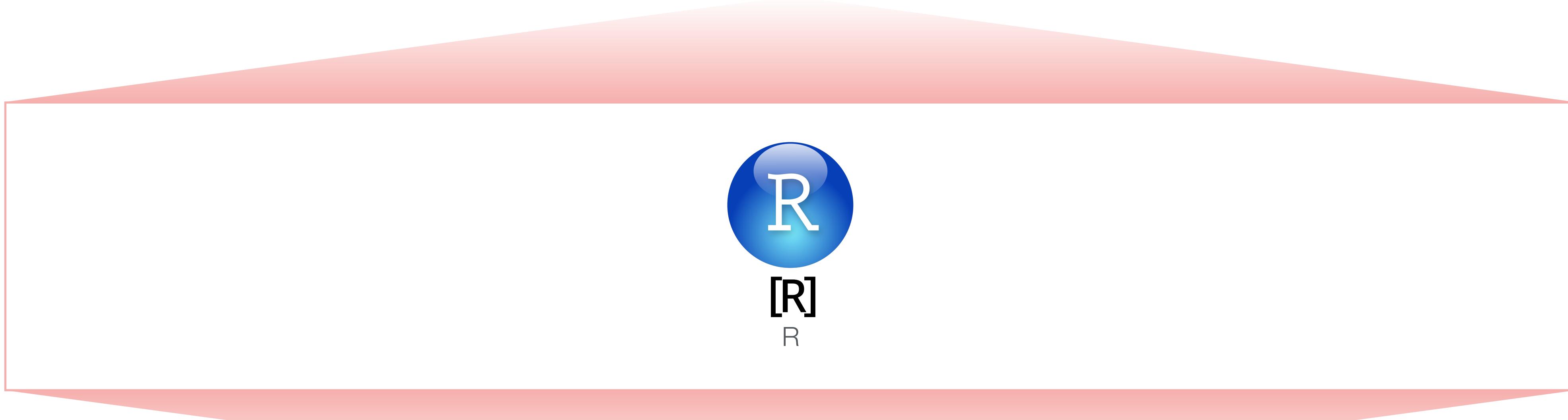
[SPSS]
SPSS



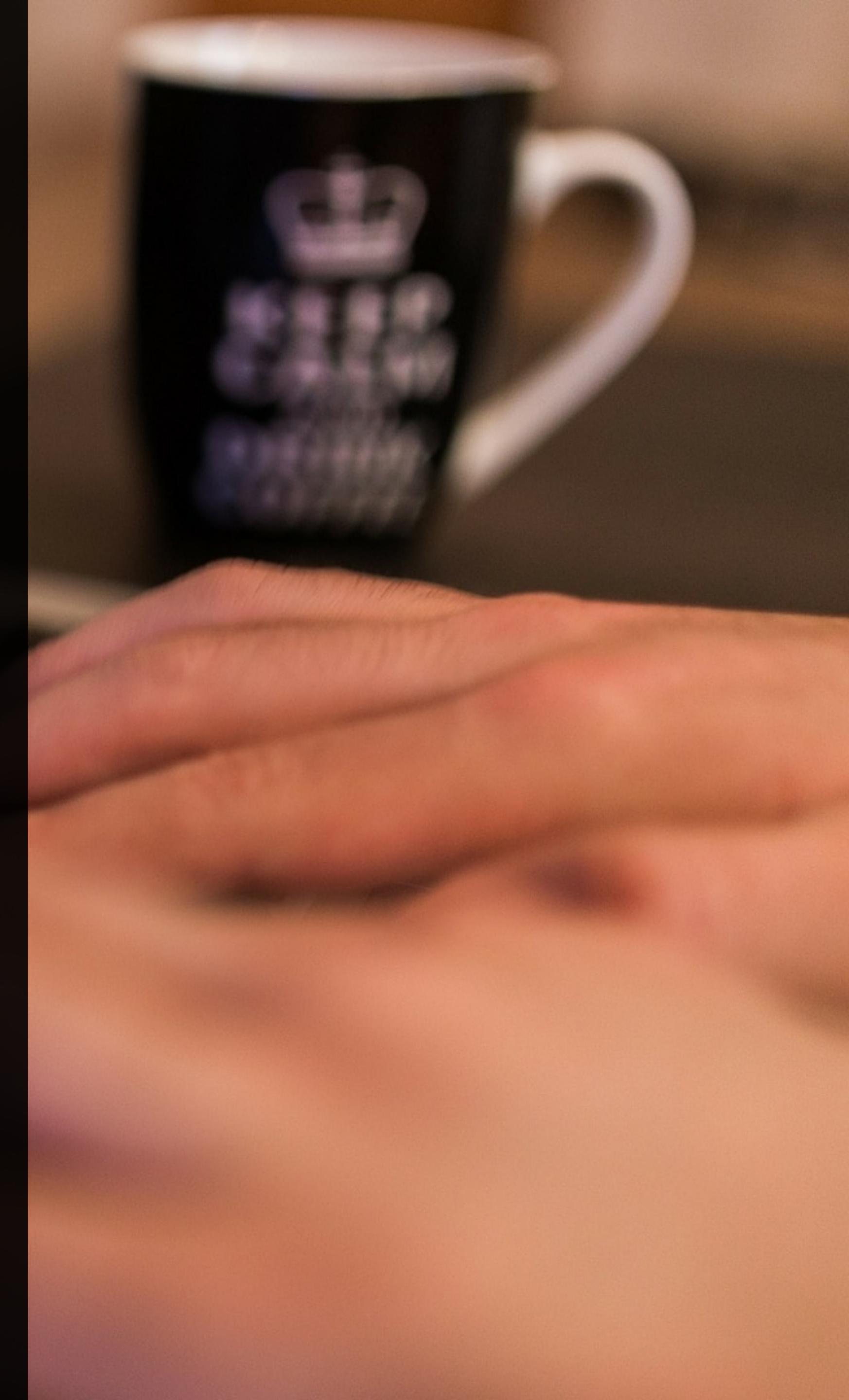
[인사이트]
INSIGHT



[데이터]
DATA

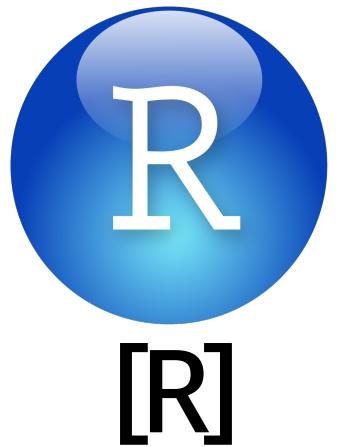


[인사이트]
INSIGHT





WHAT IS R?



WHAT IS R?

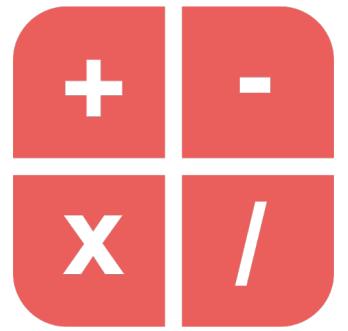
R is a **programming language** and software environment for **statistical computing** and **graphics**.

— Wikipedia



[R]

WHAT IS R?



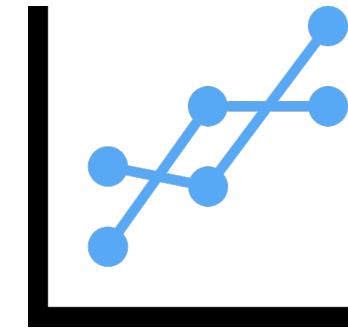
[계산기]

CALCULATOR



[프로그래밍 언어]

PROGRAMMING LANGUAGE

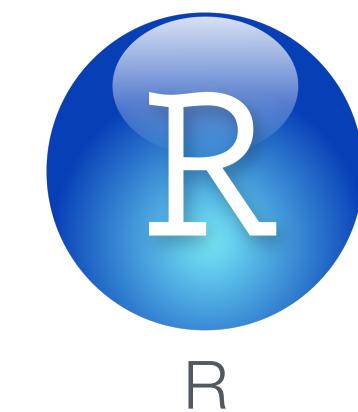


[그래픽 도구]

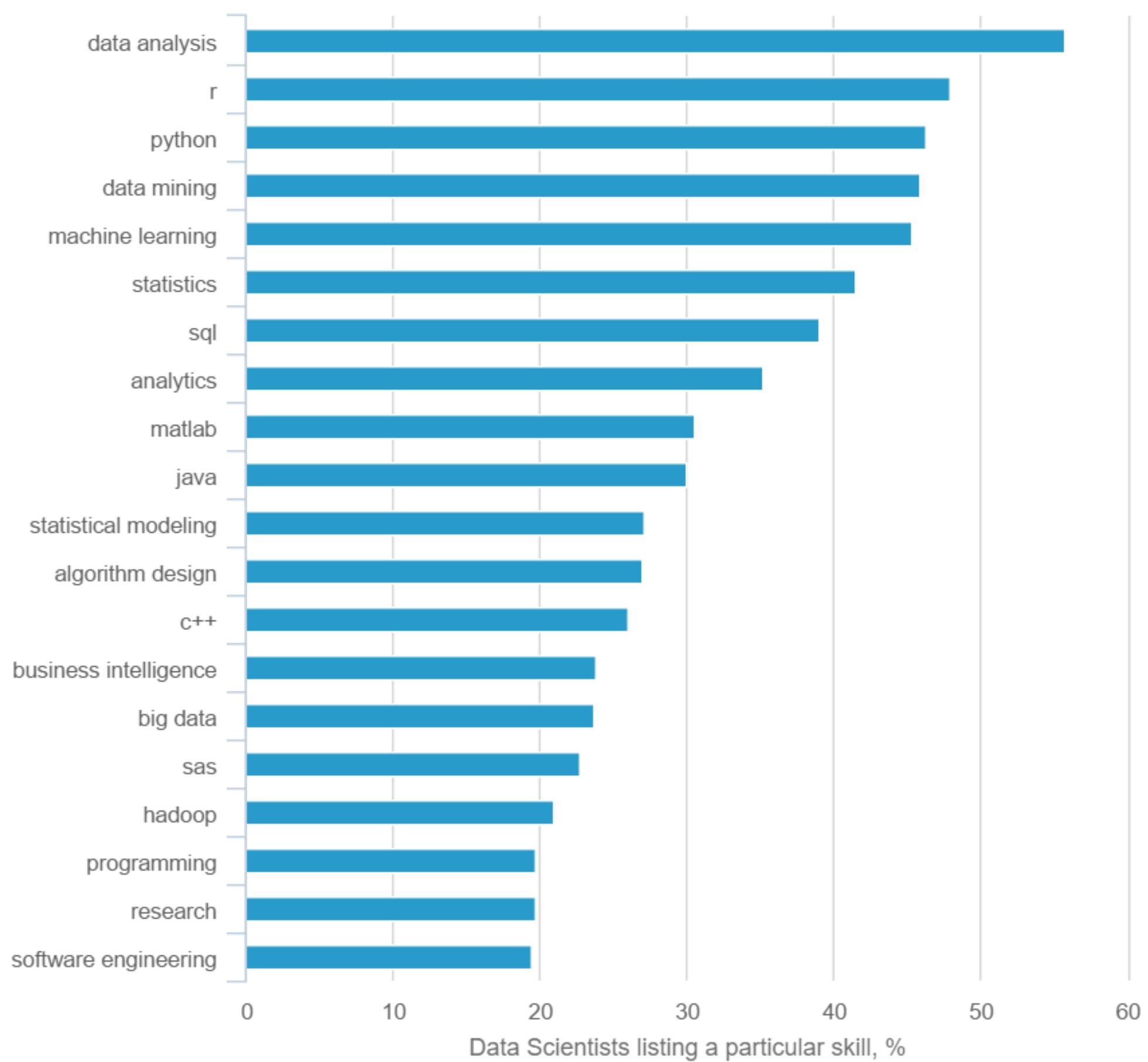
GRAPHICAL TOOL



WHY R?

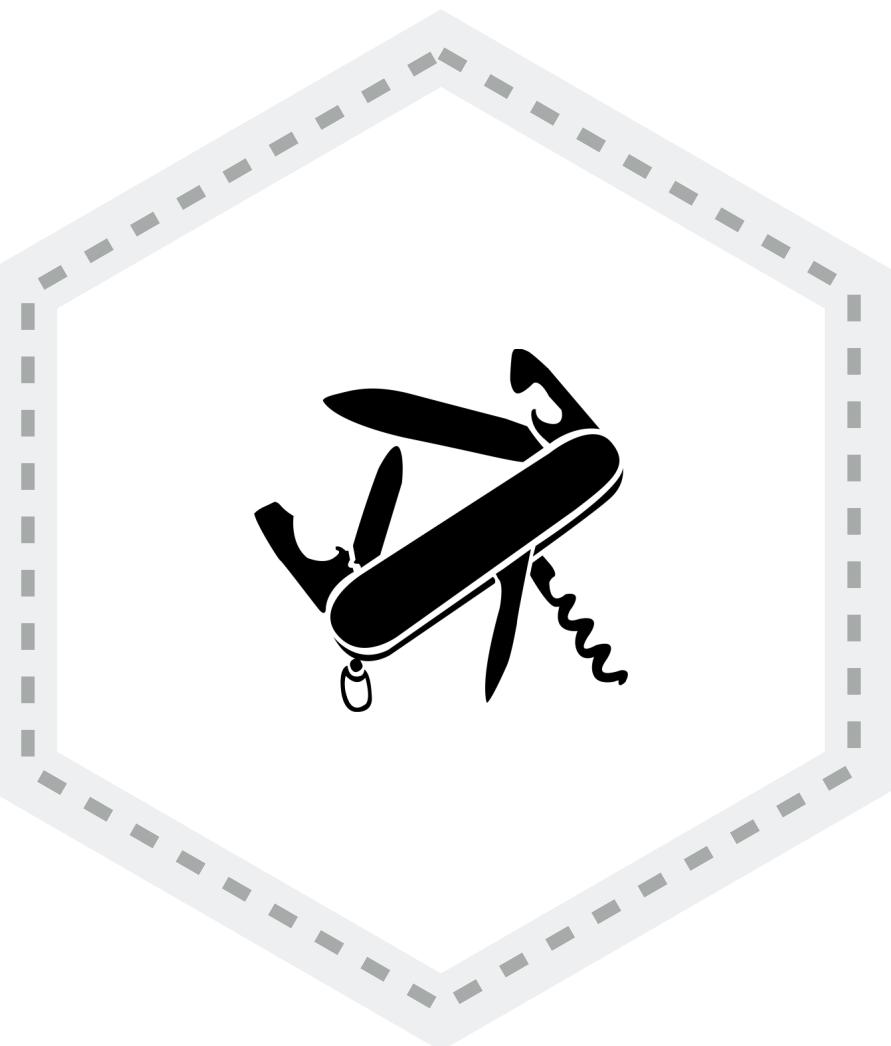


TOP 20 SKILLS OF A DATA SCIENTIST

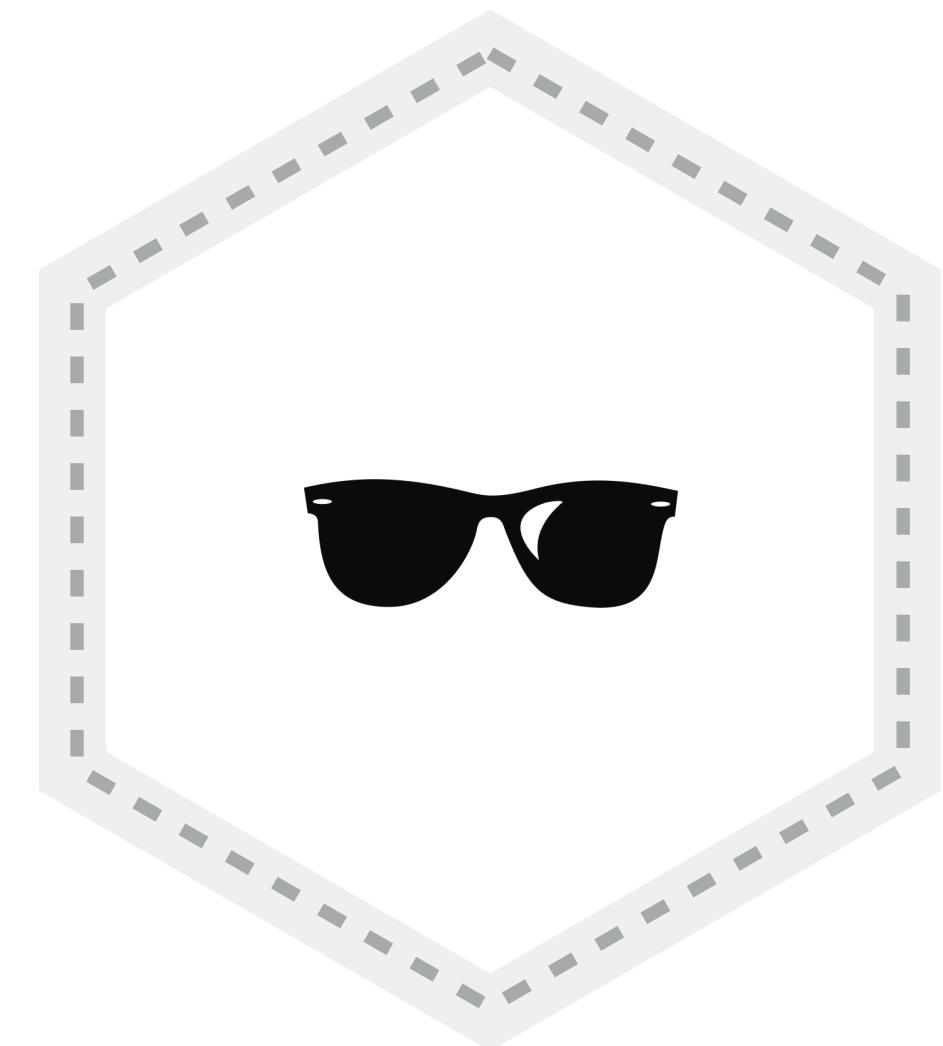




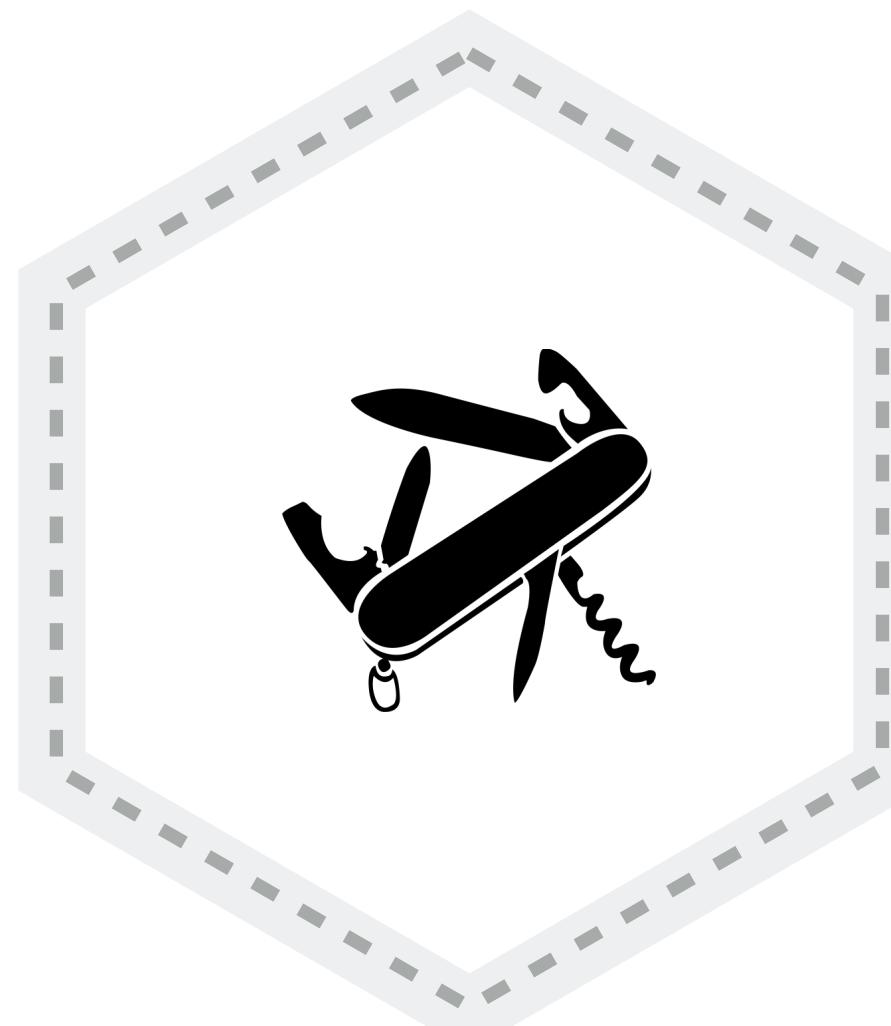
Open Source



Versatile



Trendy



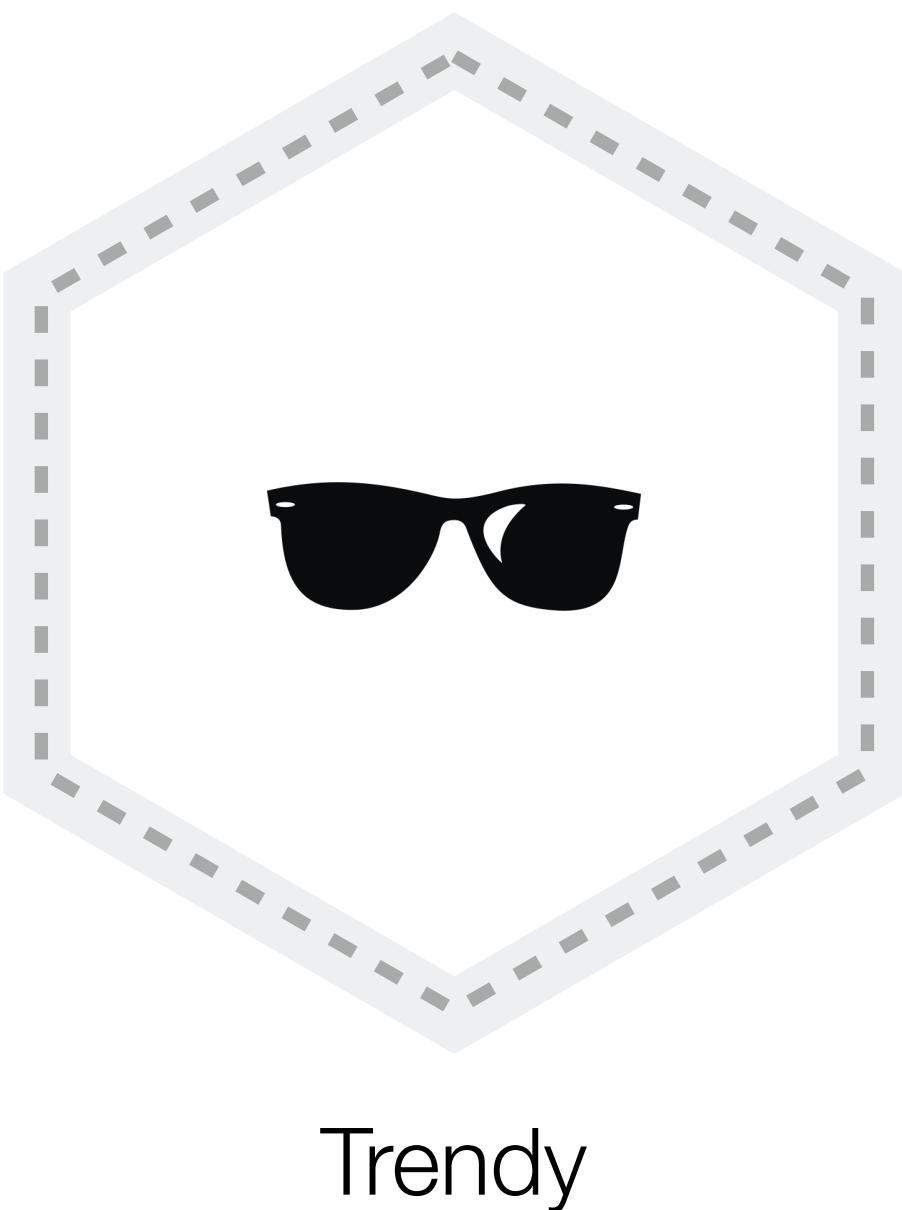
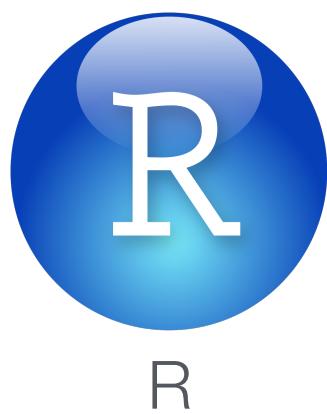
Versatile

Data Handling

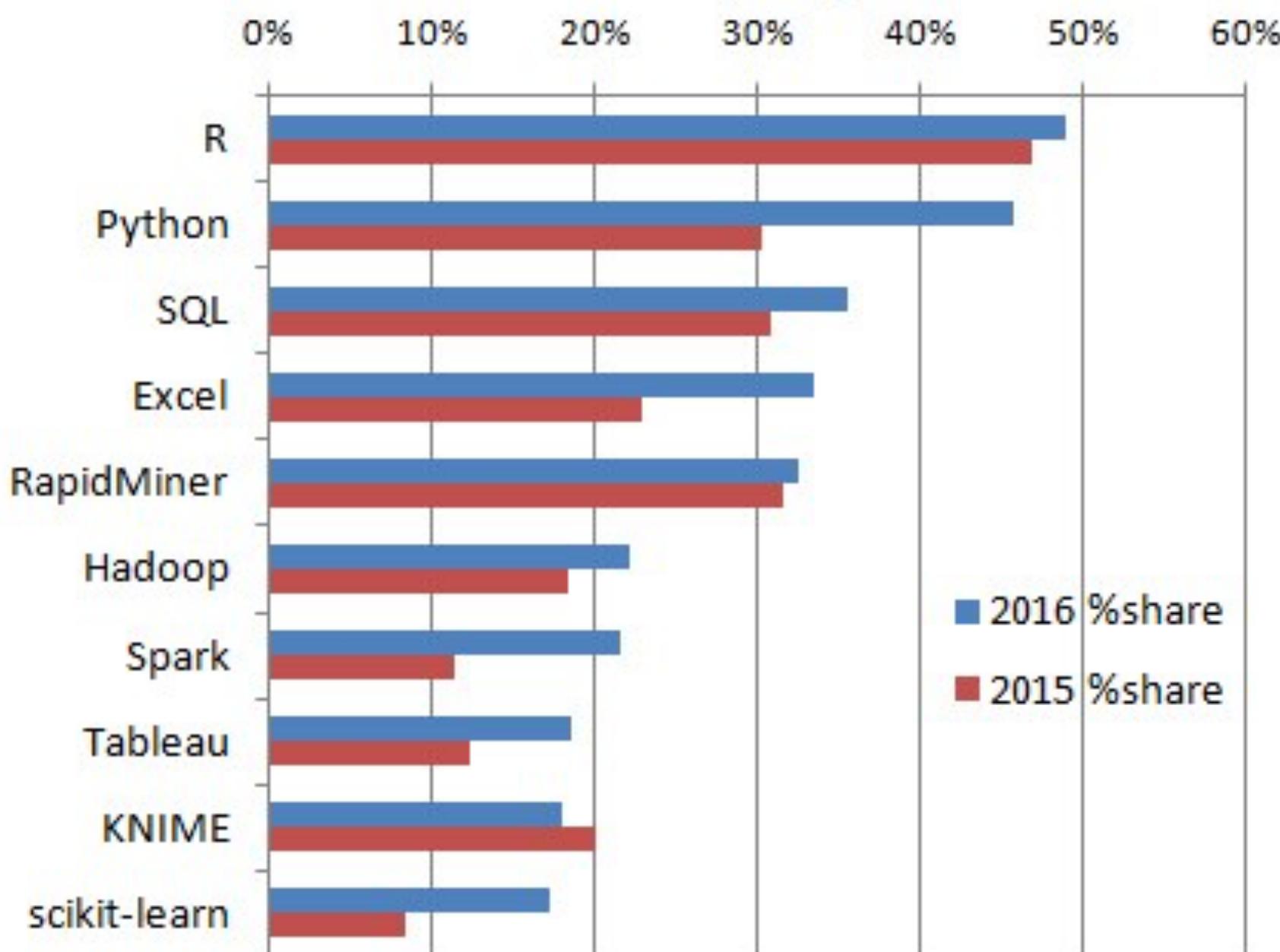
Data Visualization

Machine Learning

Reproducible Research

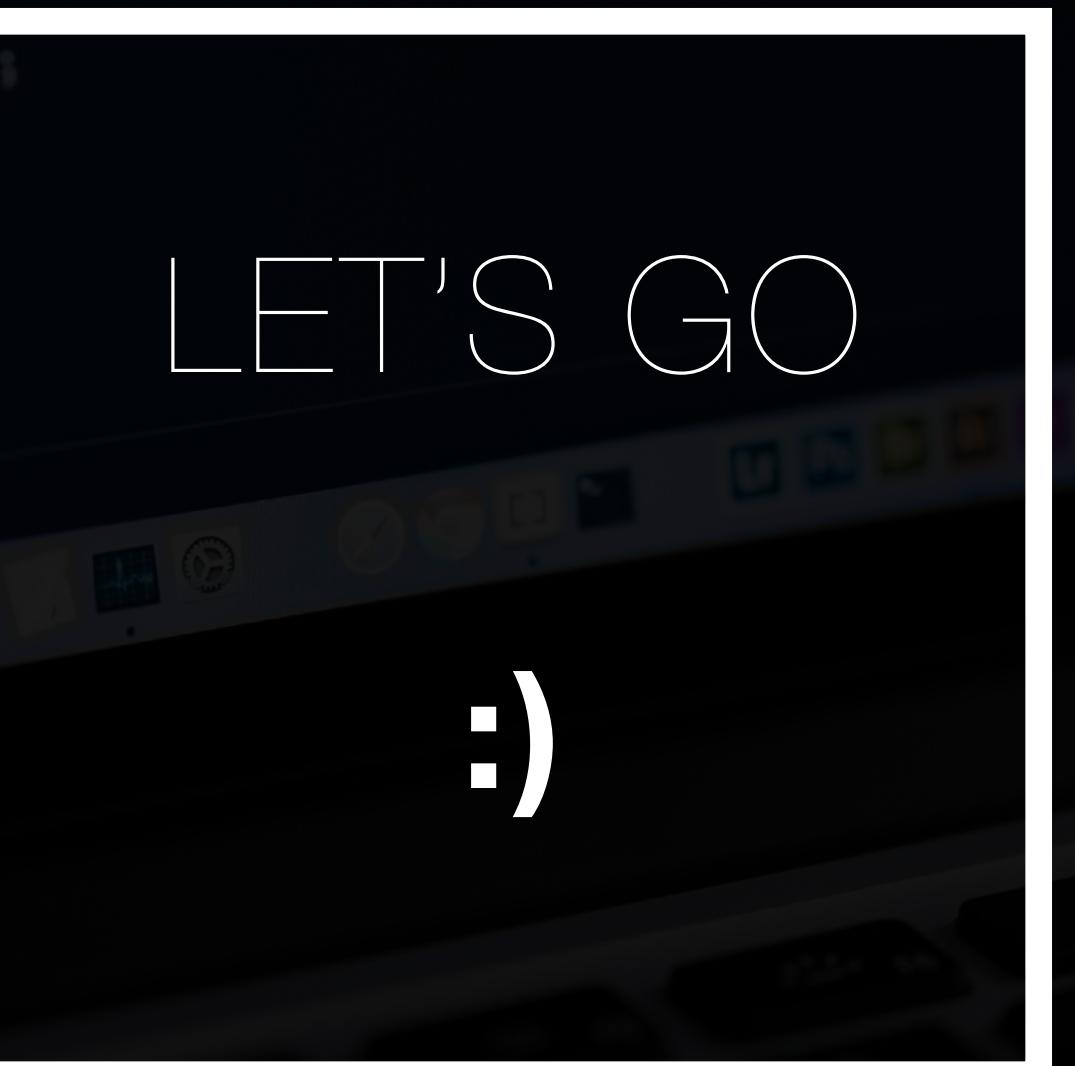


KDnuggets Analytics/Data Science 2016 Software Poll, top 10 tools





```
if ( ( $loop + 1 ) % $columns == 0 )
    $classes[] = 'last';
$image_link = wp_get_attachment_url( $attachment_id );
if ( ! $image_link )
    continue;
$image = wp_get_attachment_image( $attachment_id, apply_filters(
    'shop_thumbnail_size', 'shop_single' ) );
$image_class = esc_attr( implode( ' ', $classes ) );
$image_title = esc_attr( get_the_title( $attachment_id ) );
printf( ' <div class="slide easyzoom"><a href="%s" title="%s" ><img alt="'
    . $image_title . '" class="wp-post-image" /></a></div>', wp_get_attachment_url( $attachment_id ), $image_title );
```





things you should know



THINGS YOU SHOULD KNOW



[변수]

VARIABLE

프로그램에서 사용할 정보를 메모리에 저장할 때 사용하는 이름



THINGS YOU SHOULD KNOW



[변수]

VARIABLE

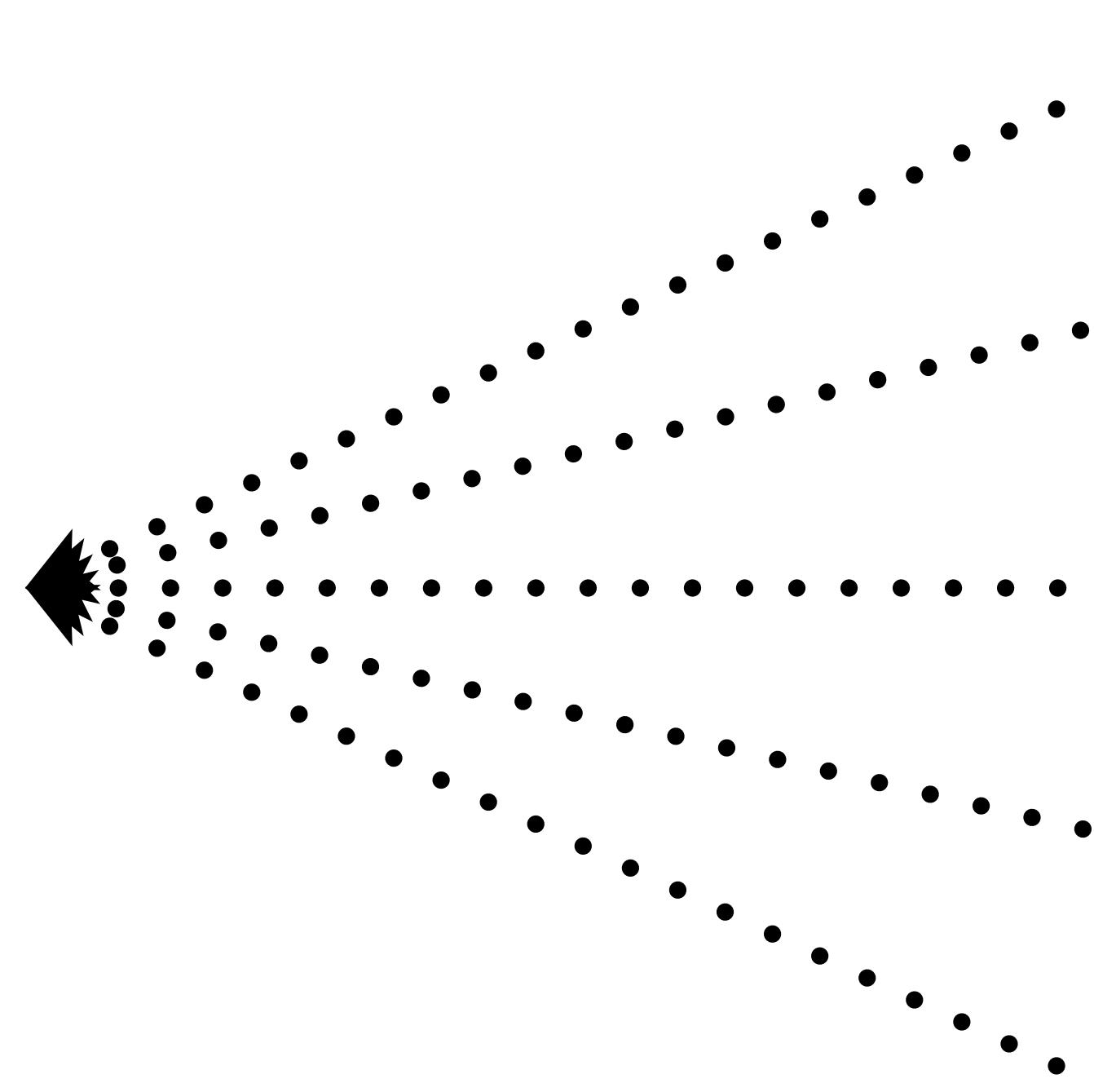
NUMBERS

STRING

DATA

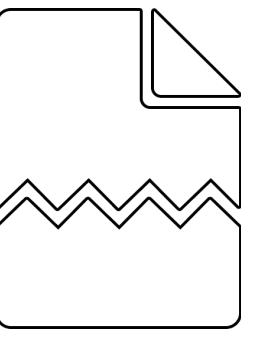
GRAPHICS

MODEL





THINGS YOU SHOULD KNOW



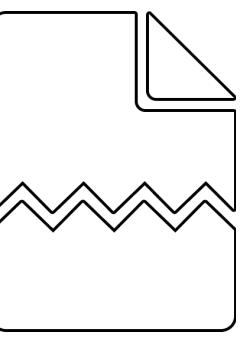
[결측값]

MISSING VALUES

데이터에서 알 수 없는 값



THINGS YOU SHOULD KNOW



[결측값]

MISSING VALUES

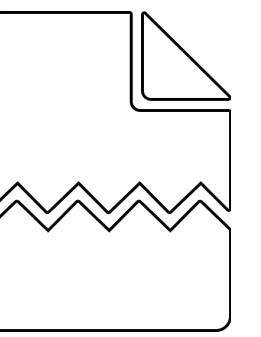
Not **A**vailable

Not **a** **N**umber

NULL



THINGS YOU SHOULD KNOW



[결측값]

MISSING VALUES

NA

NaN

NULL



THINGS YOU SHOULD KNOW



[데이터 타입]

DATA TYPE



THINGS YOU SHOULD KNOW



DATA TYPE

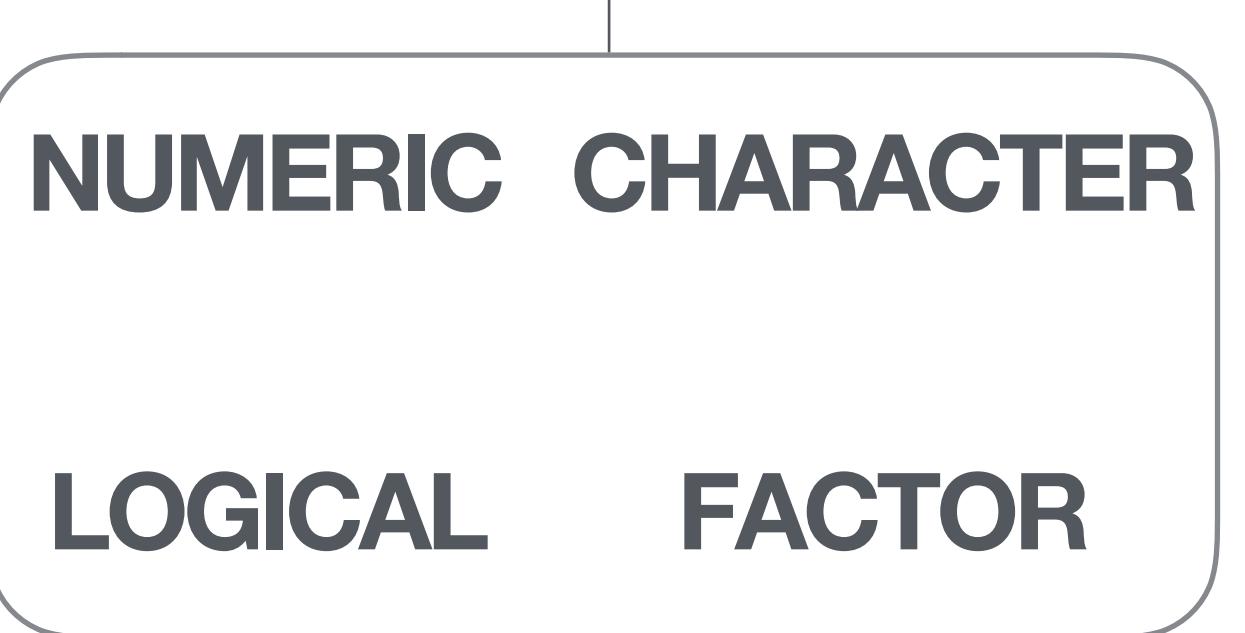
VECTOR

MATRIX

LIST

DATA.FRAME

ARRAY





VECTOR

같은 타입의 스칼라 값을 묶어 놓은 데이터 타입



DATA TYPE

MATRIX

2차원 벡터



LIST

name

A, B ,C

score

90, 80, 70

grade

A+, B+, B0



DATA.FRAME

여러분이 엑셀로 다루는 그 데이터

제어문

CONTROL
STATEMENT

```
if ( ( $loop + 1 ) % $columns == 0 )
    $classes[] = 'last';

$image_link = wp_get_attachment_url( $attachment_id );

if ( ! $image_link )
    continue;

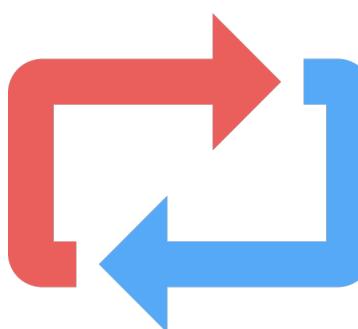
$image    = wp_get_attachment_image( $attachment_id, apply_filters(
    'shop_single_product_thumbnail_size', 'shop_single' ) );
$image_class = esc_attr( implode( ' ', $classes ) );
$image_title = esc_attr( get_the_title( $attachment_id ) );
printf( ' <div class="slide easyzoom"><a href="%s" title="%s" ><img alt="%s" /></a></div>', wp_get_attachment_url( $attachment_id ), $image_title, $image );
```



[제어문]

CONTROL STATEMENT

프로그래밍 언어에서 실행 순서를 변경시키거나 조건에 따라 실행해야 하는 명령문을 제어하는 문법

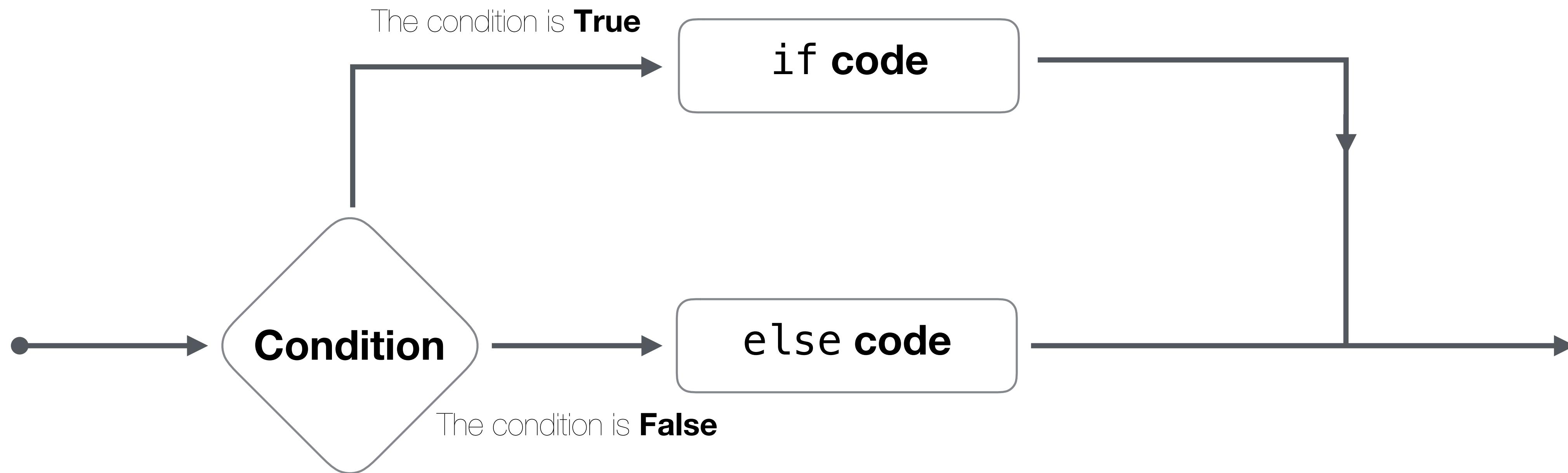


[제어문]

CONTROL STATEMENT

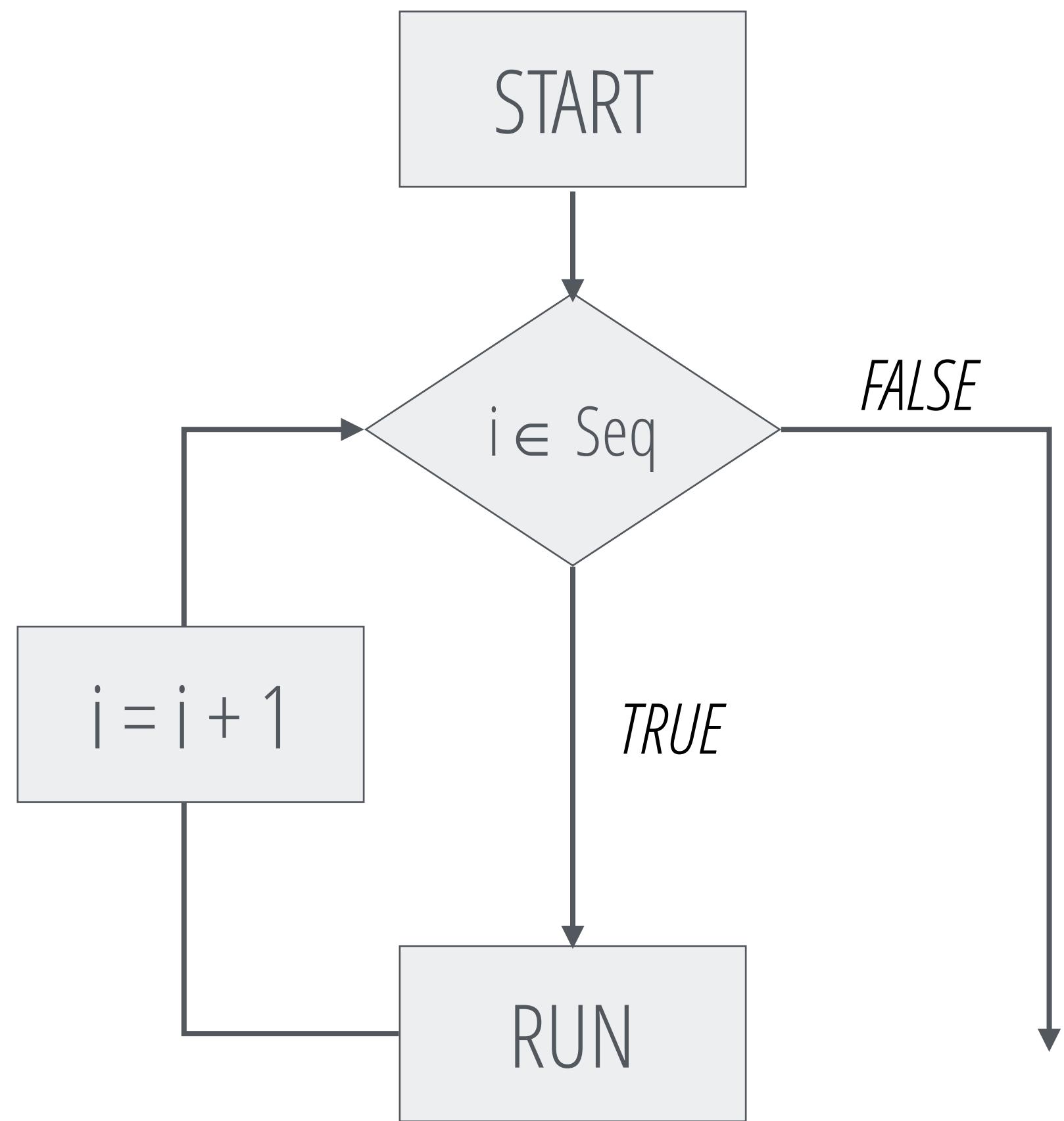
IF - ELSE

조건문



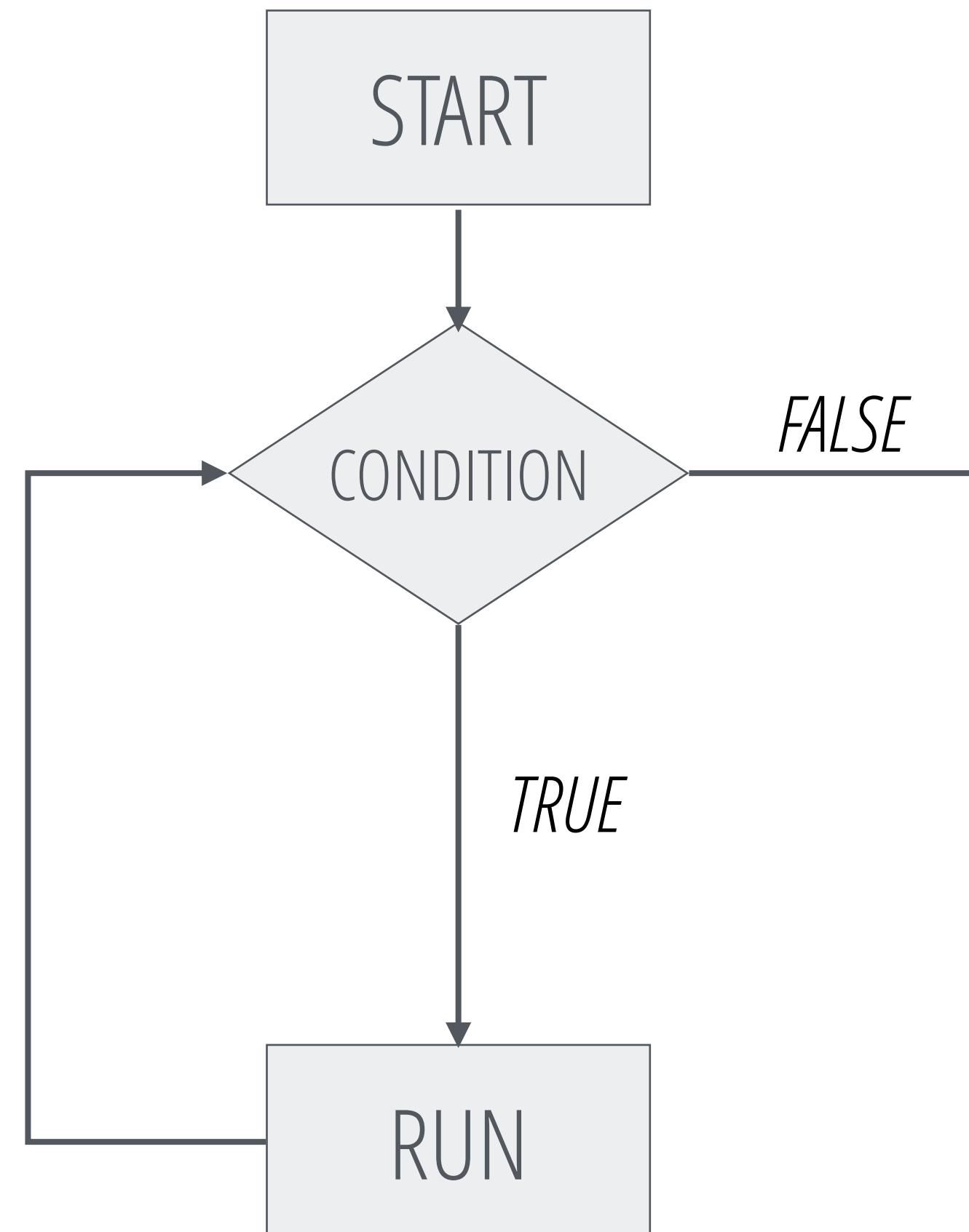


CONTROL STATEMENT



FOR

반복문



WHILE

반복문



조건문과 반복문을 이용해서
구구단을 출력해보자



과제

구구단 출력

조건문과 반복문을 이용해서
구구단을 출력해보자

$2 * 1 = 2$	$2 * 2 = 4$	$2 * 3 = 6$	$2 * 4 = 8$	$2 * 5 = 10$
$2 * 6 = 12$	$2 * 7 = 14$	$2 * 8 = 16$	$2 * 9 = 18$	$2 * 10 = 20$
=====				
$3 * 1 = 3$	$3 * 2 = 6$	$3 * 3 = 9$	$3 * 4 = 12$	$3 * 5 = 15$
$3 * 6 = 18$	$3 * 7 = 21$	$3 * 8 = 24$	$3 * 9 = 27$	$3 * 10 = 30$
=====				
$4 * 1 = 4$	$4 * 2 = 8$	$4 * 3 = 12$	$4 * 4 = 16$	$4 * 5 = 20$
$4 * 6 = 24$	$4 * 7 = 28$	$4 * 8 = 32$	$4 * 9 = 36$	$4 * 10 = 40$
=====				
$5 * 1 = 5$	$5 * 2 = 10$	$5 * 3 = 15$	$5 * 4 = 20$	$5 * 5 = 25$
$5 * 6 = 30$	$5 * 7 = 35$	$5 * 8 = 40$	$5 * 9 = 45$	$5 * 10 = 50$
=====				
$6 * 1 = 6$	$6 * 2 = 12$	$6 * 3 = 18$	$6 * 4 = 24$	$6 * 5 = 30$
$6 * 6 = 36$	$6 * 7 = 42$	$6 * 8 = 48$	$6 * 9 = 54$	$6 * 10 = 60$
=====				
$7 * 1 = 7$	$7 * 2 = 14$	$7 * 3 = 21$	$7 * 4 = 28$	$7 * 5 = 35$
$7 * 6 = 42$	$7 * 7 = 49$	$7 * 8 = 56$	$7 * 9 = 63$	$7 * 10 = 70$
=====				
$8 * 1 = 8$	$8 * 2 = 16$	$8 * 3 = 24$	$8 * 4 = 32$	$8 * 5 = 40$
$8 * 6 = 48$	$8 * 7 = 56$	$8 * 8 = 64$	$8 * 9 = 72$	$8 * 10 = 80$
=====				
$9 * 1 = 9$	$9 * 2 = 18$	$9 * 3 = 27$	$9 * 4 = 36$	$9 * 5 = 45$
$9 * 6 = 54$	$9 * 7 = 63$	$9 * 8 = 72$	$9 * 9 = 81$	$9 * 10 = 90$
=====				

이중 FOR 문과 조건문을 이용한다.
한 줄에는 최대 다섯개를 출력하고
한 단이 끝나면 ===로 구분을 해준다.

과제

구구단 출력

조건문과 반복문을 이용해서
구구단을 출력해보자

사용하면 좋은 함수들 + 힌트

cat()

$x * y = xy$ 에서 y 가 5가 되면 한 줄을 띄면 된다.

$x * y = xy$ 에서 y 가 100이 되면?

\n 과 \t 는 각각 무엇일까?

THX :)