R

GETTING STARTED WITH R

KHU **SNS**
HAN JAEYOON

WHO AM I

HAN JAEYOON

Dept. of SNS, Master's Course

Kaggler

R User

WHO AM I

Knowledge • 233 teams

Kobe Bryant Shot Selection

Fri 15 Apr 2016 — Mon 13 Jun 2016 (39 days to go)

Dashboard

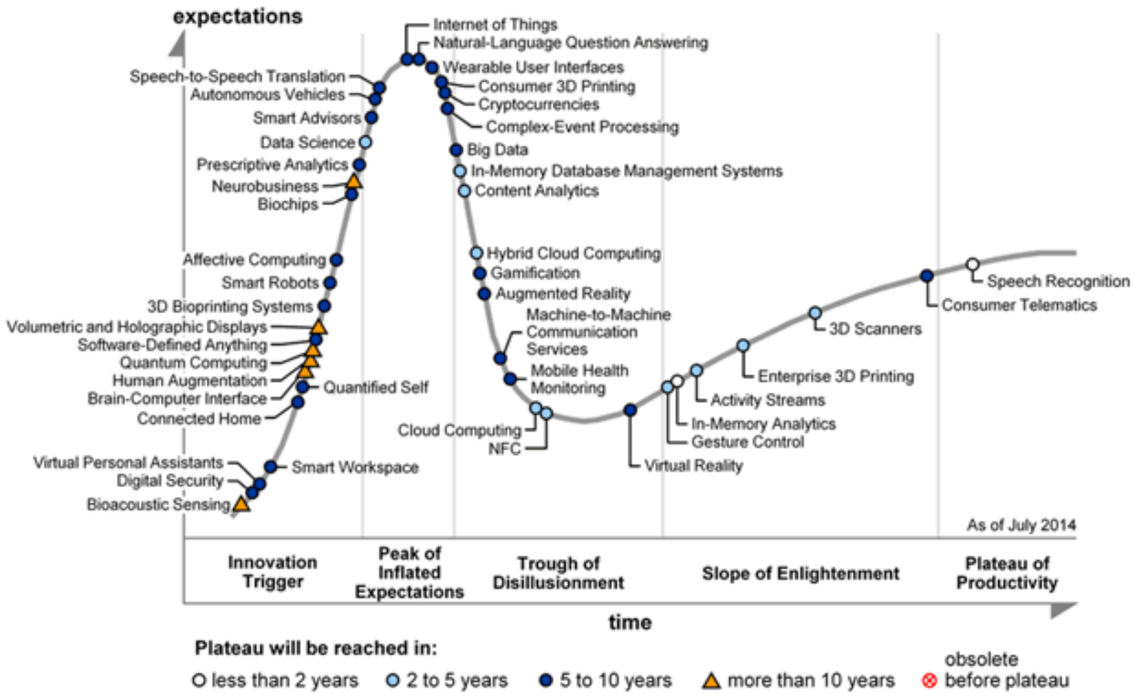
Public Leaderboard - Kobe Bryant Shot Selection

#	Δ5d	Team Name * in the money	Score	Entries	Last Submission UTC (Best – Last Submission)
1	new	LeBronLearnsML *	0.59509	5	Mon, 02 May 2016 01:46:12
2	↑13	Dagar Katyal	0.59908	13	Wed, 04 May 2016 00:31:40 (-0.5h)
3	↓2	Jason Noriega	0.59915	28	Thu, 21 Apr 2016 04:13:59 (-26.9h)
4	↓2	Josh Stone	0.59934	1	Tue, 19 Apr 2016 18:24:57
5	↓2	Juan Aguilera	0.59953	5	Thu, 28 Apr 2016 02:11:47 (-9.2h)
6	↓2	anokas	0.60014	10	Thu, 28 Apr 2016 10:14:16 (-6.6d)
7	↓2	Jeff Mills	0.60024	48	Thu, 28 Apr 2016 09:25:35 (-10h)
8	**↓1**	**Jae-YoonHan**	**0.60064**	**56**	**Thu, 05 May 2016 13:50:21**

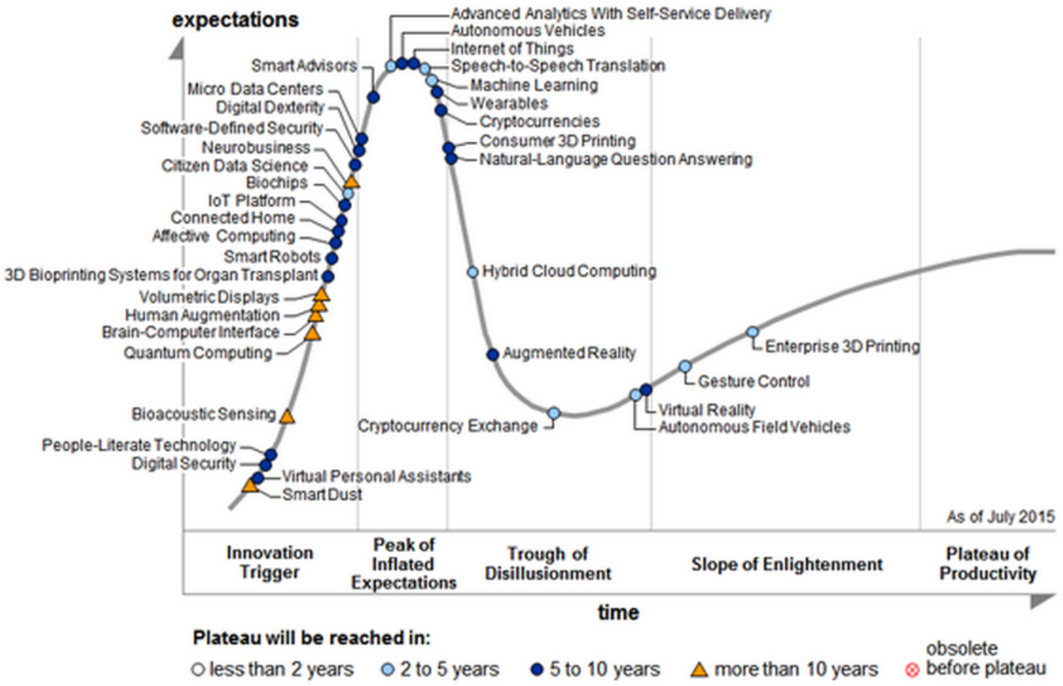
BIG DATA?

BIG DATA?

GARTNER HYPE CYCLE



2014



2015

THE EMERGENCE OF BIG DATA

ESSENCE

DATA ANALYSIS
ALREADY IMPLEMENTED

THEN?

DATA SCIENCE

DATA SCIENCE

= 데이터에서 + 패턴을 찾아내어 + 비즈니스 기회로

—하용호, 넘버웍스 대표

THE MORE, THE BETTER...?

2TB

= 2,048 GB

→ **BIG DATA**

CAN YOU HANDLE IT?

QUOTES

> *If you cannot do great things,*
> *do small things in a great way.*
>
> — Napoleon Hill

WHAT IS R?

PROGRAMMING LANGUAGE

STATISTICAL COMPUTING + GRAPHICS

DATA ANALYSIS

WHY R?

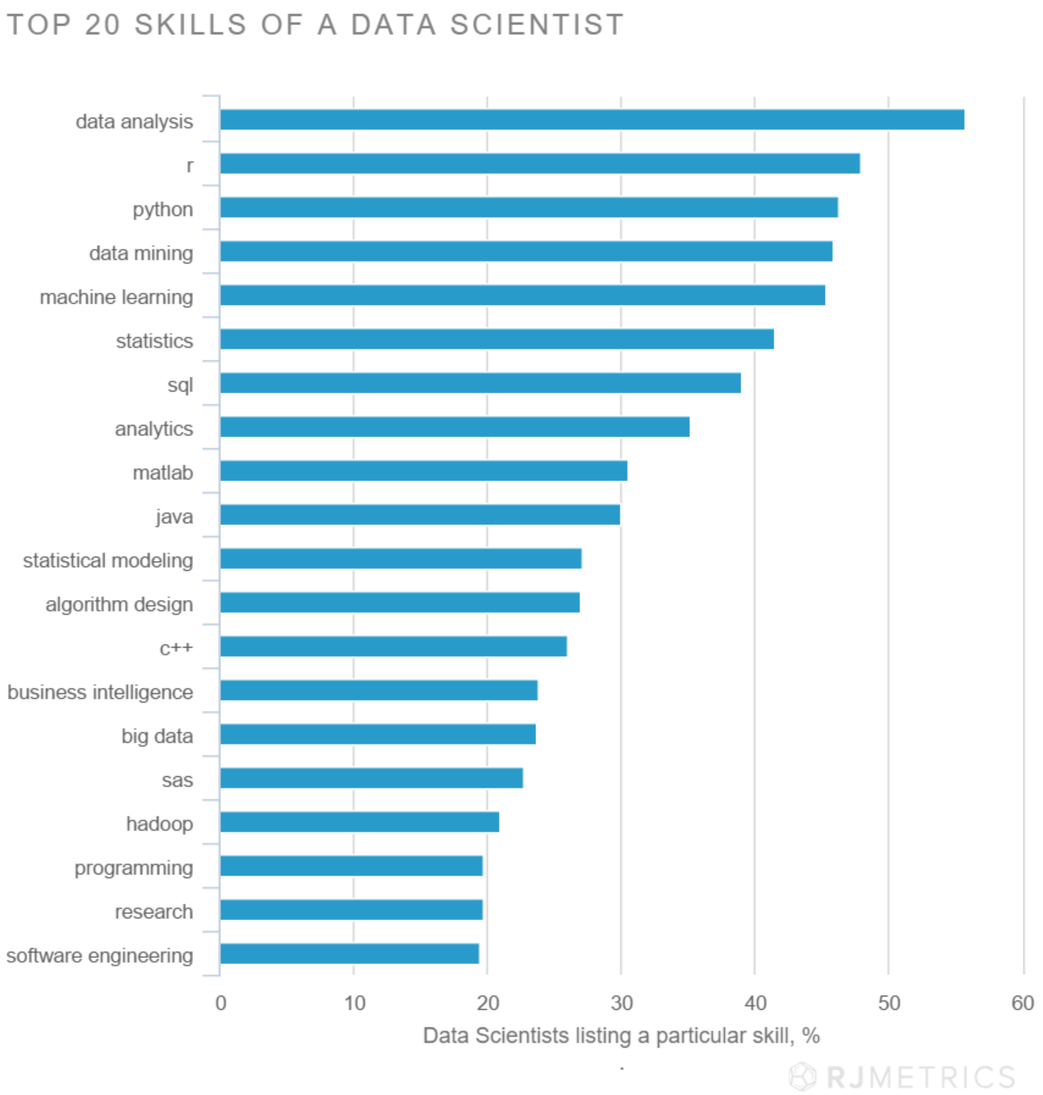
FREE

NUMEROUS LIBRARIES FOR DATA ANALYSIS

EASY TO LEARN

A LOT OF STUDY MATERIAL

WHY R?


TOP 20 SKILLS OF A DATA SCIENTIST
data analysis
r
python
data mining
machine learning
statistics
sql
analytics
matlab
java
statistical modeling
algorithm design
c++
business intelligence
big data
sas
hadoop
programming
research
software engineering
0
10
20
30
40
50
60
Data Scientists listing a particular skill, %
RJMETRICS

WITH R?

OUR GOAL

BECOME A

BIG DATA EXPERT

SELF-TAUGHT R USER

OUR GOAL

SELF-TAUGHT R USER

CURRICULUM

WEEK 1 HOW TO USE RSTUDIO
R BASIC SYNTAX
DATA TYPE

WEEK 2 CONTROL STATEMENTS
DATA MANIPULATION I

WEEK 3 DATA MANIPULATION II
DATA VISUALIZATION I

WEEK 4 DATA VISUALIZATION II
MACHINE LEARNING I

WEEK 5 MACHINE LEARNING II
FINALE