

The 3 Views of Data

Stat 133 by Gaston Sanchez

Creative Commons Attribution Share-Alike 4.0 International CC BY-SA

Last time ...

Last time

- We talked about the RQs
- We saw a toy example: Points and Salary
- We carried out a (very informal) DAC
- We introduced the idea of looking at “Data”
from three different perspectives

The Data Computing Diagram (DCD)

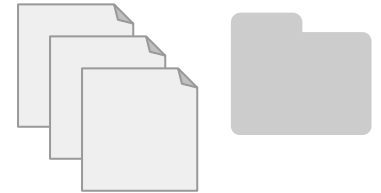
Data
Sets



Software &
Languages



Code, Scripts,
Programs



Computers



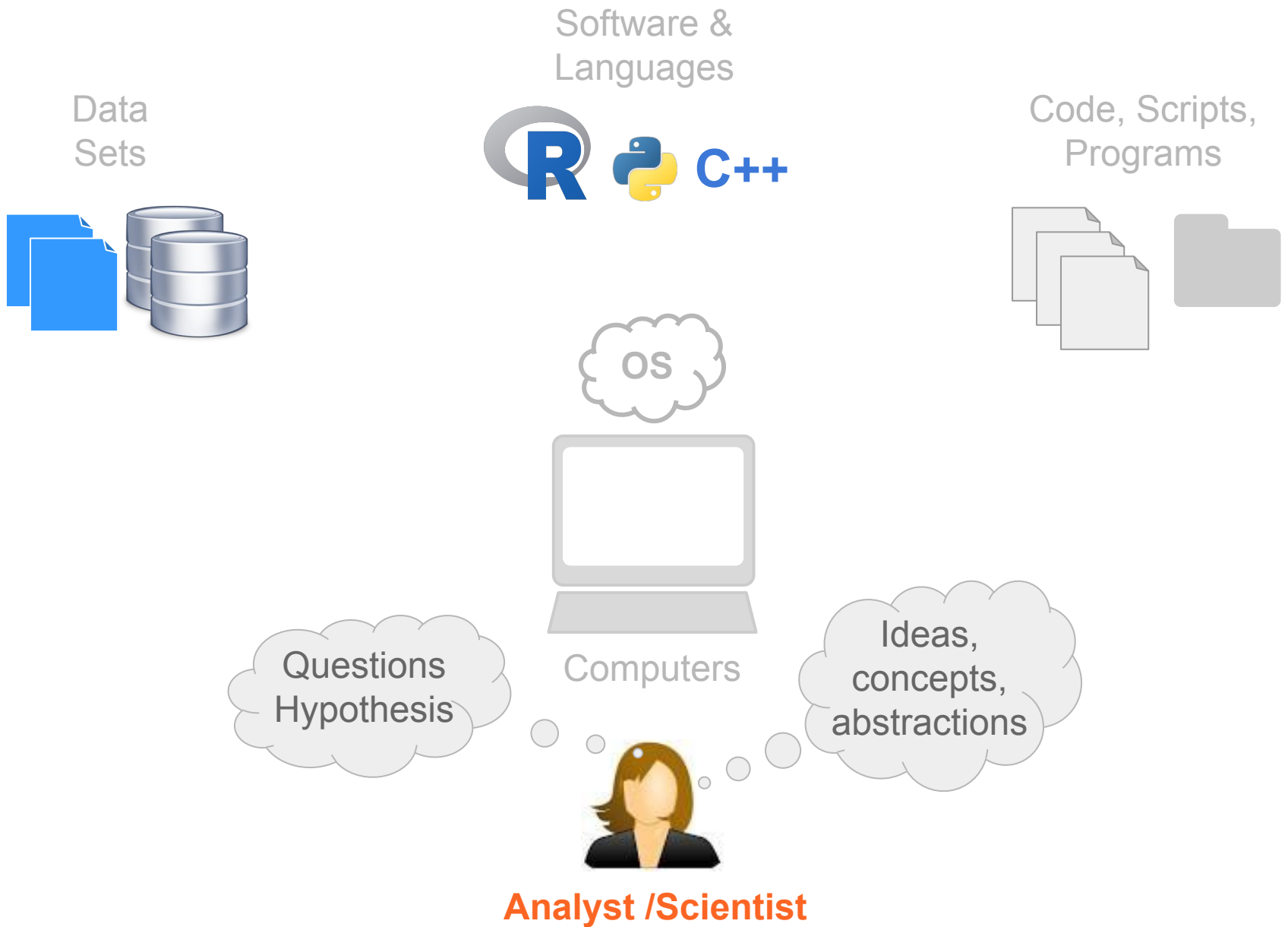
Analyst /Scientist

Research Question



The more scored **points**,
the higher the **salary**?

How do
statisticians/scientists
think about data?



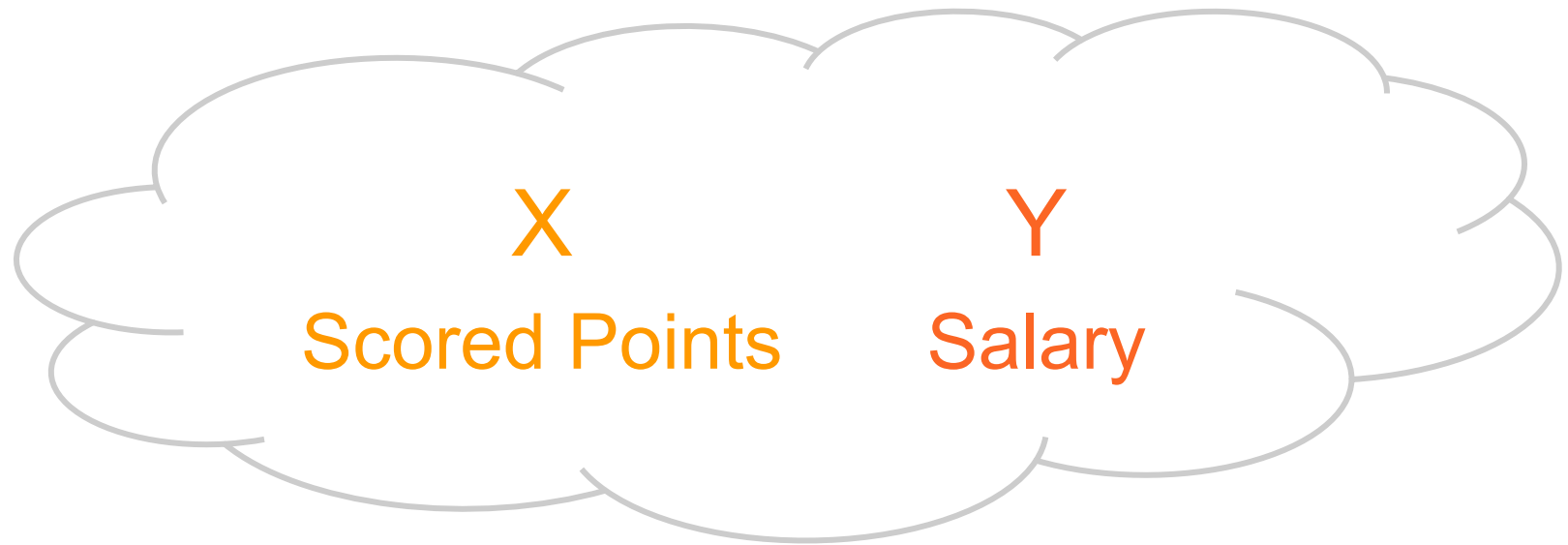
Research Question #1

The more scored **points**,
the higher the **salary**?



Analyst /Scientist

Statistical Perspective



Quantitative
variables



Analyst /Scientist

Statistical Perspective

`r = cor(Salary, Points)`

Summaries



Analyst /Scientist

Statistical Perspective

$$Y = f(X) + e$$

$$\text{Salary} = f(\text{Points}) + e$$

Theoretical
Model



Analyst /Scientist

Statistical Perspective

$$Y = b_0 + b_1 X + e$$

$$\text{Salary} = b_0 + b_1 \text{Points} + e$$

Linear
Model



Analyst /Scientist

Data (as in Statistics)

Variables observed on some individuals

Variable: characteristic, feature, descriptor

Individual: objects, animals, humans, etc

Typical representation in tabular format

Referring to variables in statistics ...

Quantitative -vs- Qualitative

Continuous -vs- Discrete

Numerical -vs- Categorical

Scales: Ratio, Interval, Ordinal, Nominal

Dependent -vs- Independent

Descriptors (predictors) -vs- Response

Input -vs- Output

Missing values, Censored

What about the Data?

NBA Data

basketball-reference.com

Roster

Share & more ▼

Glossary

No.		Pos	Ht	Wt	Birth Date		Exp	College
30	Stephen Curry	PG	6-3	190	March 14, 1988		7	Davidson College
11	Klay Thompson	SG	6-7	215	February 8, 1990		5	Washington State University
9	Andre Iguodala	SF	6-6	215	January 28, 1984		12	University of Arizona
34	Shaun Livingston	PG	6-7	192	September 11, 1985		11	
5	Kevon Looney	C	6-9	220	February 6, 1996		1	University of California, Los Angeles
23	Draymond Green	PF	6-7	230	March 4, 1990		4	Michigan State University
35	Kevin Durant	SF	6-9	240	September 29, 1988		9	University of Texas at Austin
3	David West	C	6-9	250	August 29, 1980		13	Xavier University
0	Patrick McCaw	SG	6-7	185	October 25, 1995		R	University of Nevada, Las Vegas
27	Zaza Pachulia	C	6-11	270	February 10, 1984		13	
15	Damian Jones	C	7-0	245	June 30, 1995		R	Vanderbilt University
20	James Michael McAdoo	PF	6-9	230	January 4, 1993		2	University of North Carolina
21	Ian Clark	SG	6-3	175	March 7, 1991		3	Belmont University
1	JaVale McGee	C	7-0	270	January 19, 1988		8	University of Nevada, Reno
22	Matt Barnes	SF	6-7	226	March 9, 1980		13	University of California, Los Angeles

Show a glossary for each term in the table below
or hold your mouse over the header

Roster

Share & more ▼

Glossary

Also view explanations by holding mouse over column headers

No. -- Uniform Number

Pos -- Position

Ht -- Height

Wt -- Weight

Exp -- Years experience in NBA/ABA (prior to this season)



College

[University of California, Los Angeles](#)

[Belmont University](#)

[Davidson College](#)

[University of Texas at Austin](#)

[Michigan State University](#)

[University of Arizona](#)

[Vanderbilt University](#)

[University of California, Los Angeles](#)

[University of North Carolina](#)

[University of Nevada, Las Vegas](#)

[University of Nevada, Reno](#)

[Washington State University](#)

[Virginia Commonwealth University](#)

[Xavier University](#)

0	Patrick McCaw	SG	6-7	185	October 25, 1995		R	University of Nevada, Las Vegas
1	JaVale McGee	C	7-0	270	January 19, 1988		8	University of Nevada, Reno
27	Zaza Pachulia	C	6-11	270	February 10, 1984		13	
11	Klay Thompson	SG	6-7	215	February 8, 1990		5	Washington State University
18	Anderson Varejao	C	6-10	273	September 28, 1982		12	
2	Briante Weber	PG	6-2	165	December 29, 1992		1	Virginia Commonwealth University
3	David West	C	6-9	250	August 29, 1980		13	Xavier University

Roster

Share & more ▼

Glossary

No.		Pos	Ht	Wt	Birth Date		Exp	College
30	Stephen Curry	PG	6-3	190	March 14, 1988		7	Davidson College
11	Klay Thompson	SG	6-7	215	February 8, 1990		5	Washington State University
9	Andre Iguodala	SF	6-6	215	January 28, 1984		12	University of Arizona
34	Shaun Livingston	PG	6-7	192	September 11, 1985		11	
5	Kevon Looney	C	6-9	220	February 6, 1996		1	University of California, Los Angeles
23	Draymond Green	PF	6-7	230	March 4, 1990		4	Michigan State University
35	Kevin Durant	SF	6-9	240	September 29, 1988		9	University of Texas at Austin
3	David West	C	6-9	250	August 29, 1980		13	Xavier University
0	Patrick McCaw	SG	6-7	185	October 25, 1995		R	University of Nevada, Las Vegas
27	Zaza Pachulia	C	6-11	270	February 10, 1984		13	
15	Damian Jones	C	7-0	245	June 30, 1995		R	Vanderbilt University
20	James Michael McAdoo	PF	6-9	230	January 4, 1993		2	University of North Carolina
21	Ian Clark	SG	6-3	175	March 7, 1991		3	Belmont University
1	JaVale McGee	C	7-0	270	January 19, 1988		8	University of Nevada, Reno
22	Matt Barnes	SF	6-7	226	March 9, 1980		13	University of California, Los Angeles

Per Game

Share & more ▼

Glossary

Rk		Age	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%
1	Klay Thompson	26	78	78	34.0	8.3	17.6	.468	3.4	8.3	.414	4.8	9.3	.516
2	Kevin Durant	28	62	62	33.4	8.9	16.5	.537	1.9	5.0	.375	7.0	11.5	.608
3	Stephen Curry	28	79	79	33.4	8.5	18.3	.468	4.1	10.0	.411	4.4	8.3	.537
4	Draymond Green	26	76	76	32.5	3.6	8.6	.418	1.1	3.5	.308	2.5	5.1	.494
5	Andre Iguodala	33	76	0	26.3	2.9	5.5	.528	0.8	2.3	.362	2.0	3.1	.651
6	Matt Barnes	36	20	5	20.5	1.9	4.5	.422	0.9	2.6	.346	1.0	1.9	.526
7	Zaza Pachulia	32	70	70	18.1	2.3	4.4	.534	0.0	0.0	.000	2.3	4.4	.538
8	Shaun Livingston	31	76	3	17.7	2.3	4.2	.547	0.0	0.0	.333	2.3	4.1	.550
9	Patrick McCaw	21	71	20	15.1	1.5	3.5	.433	0.6	1.7	.333	0.9	1.7	.533
10	Ian Clark	25	77	0	14.8	2.7	5.6	.487	0.8	2.1	.374	1.9	3.5	.556
11	David West	36	68	0	12.6	2.0	3.7	.536	0.0	0.1	.375	1.9	3.6	.541
12	JaVale McGee	29	77	10	9.6	2.7	4.1	.652	0.0	0.1	.000	2.7	4.1	.660
13	James Michael McAdoo	24	52	2	8.8	1.2	2.3	.530	0.0	0.2	.250	1.2	2.1	.550
14	Damian Jones	21	10	0	8.5	0.8	1.6	.500	0.0	0.0		0.8	1.6	.500
15	Kevon Looney	20	53	4	8.4	1.1	2.0	.523	0.0	0.2	.222	1.0	1.8	.551
16	Anderson Varejao	34	14	1	6.6	0.4	1.0	.357	0.0	0.0		0.4	1.0	.357
17	Briante Weber	24	7	0	6.6	0.7	2.0	.357	0.0	0.4	.000	0.7	1.6	.455

Salaries

Rk		Salary
1	Kevin Durant	\$26,540,100
2	Klay Thompson	\$16,663,575
3	Draymond Green	\$15,330,435
4	Stephen Curry	\$12,112,359
5	Andre Iguodala	\$11,131,368
6	Shaun Livingston	\$5,782,450
7	Zaza Pachulia	\$2,898,000
8	David West	\$1,551,659
9	Anderson Varejao	\$1,551,659
10	JaVale McGee	\$1,403,611
11	Kevon Looney	\$1,182,840
12	Damian Jones	\$1,171,560
13	Ian Clark	\$1,015,696
14	James Michael McAdoo	\$980,431
15	Jason Thompson	\$945,126
16	Patrick McCaw	\$543,471
17	Matt Barnes	\$383,351
18	Elliot Williams	\$250,000
19	Briante Weber	\$102,898
20	Elgin Cook	\$50,000
21	Cameron Jones	\$50,000
22	Scott Wood	\$50,000
23	Phil Pressey	\$35,000

Data Collection

Do Data need to be collected?

- Designing Experiments
- Designing Surveys
- Sampling Design
- Recording & Gathering

We will assume that ...

Data is already in digital form

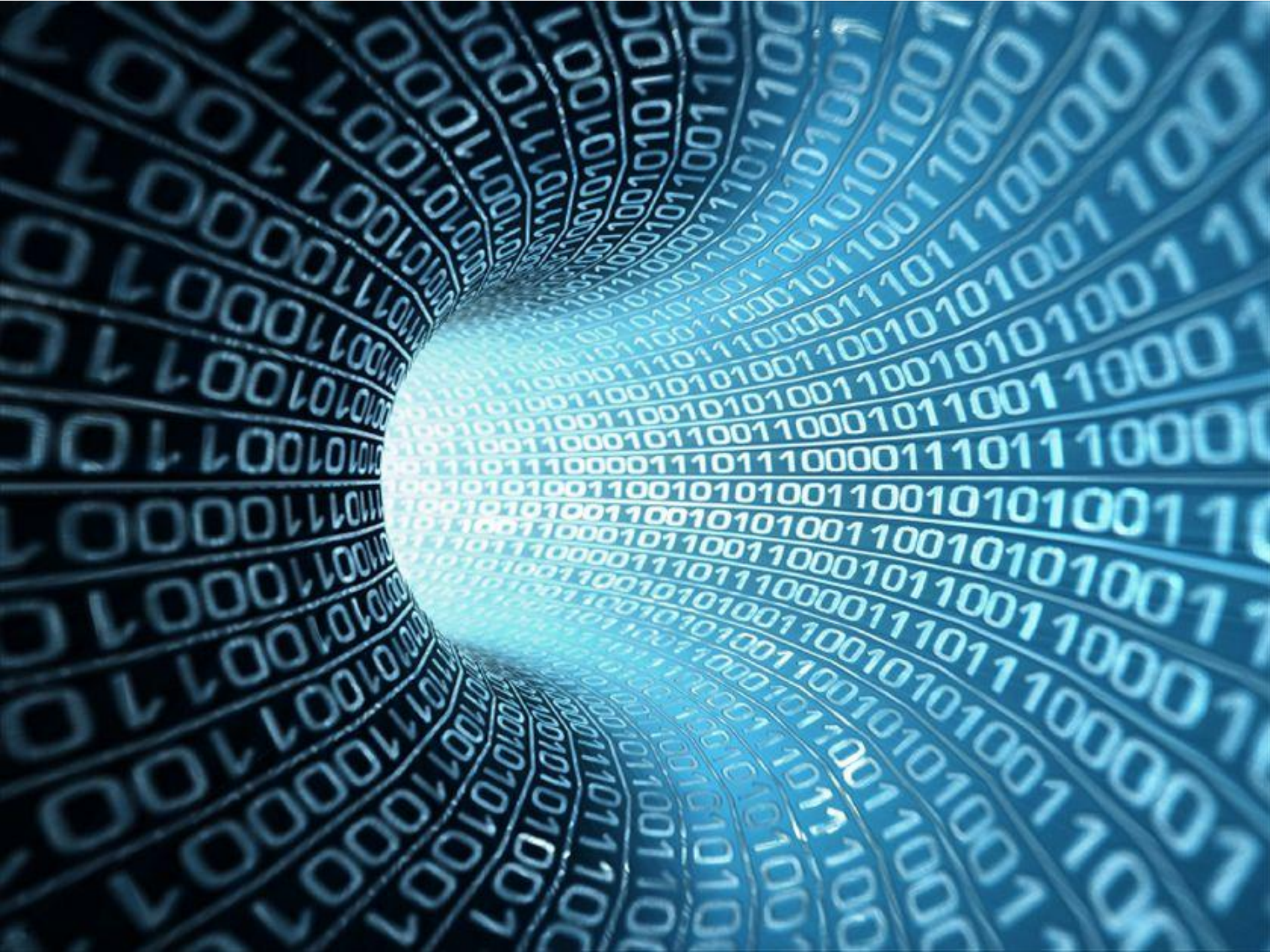
It has already been collected

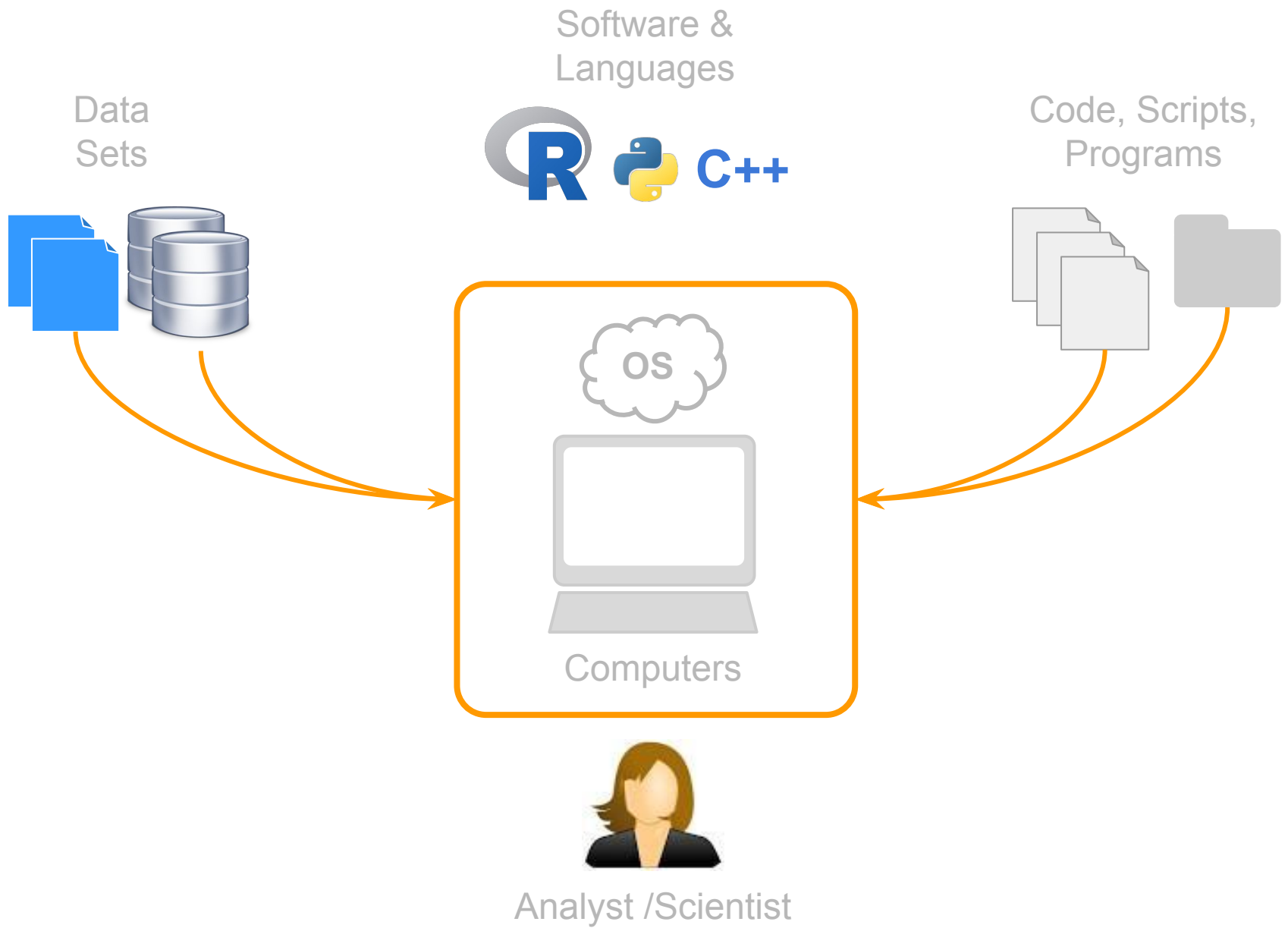
It is already stored in some files/directories

No worries about transcribing data

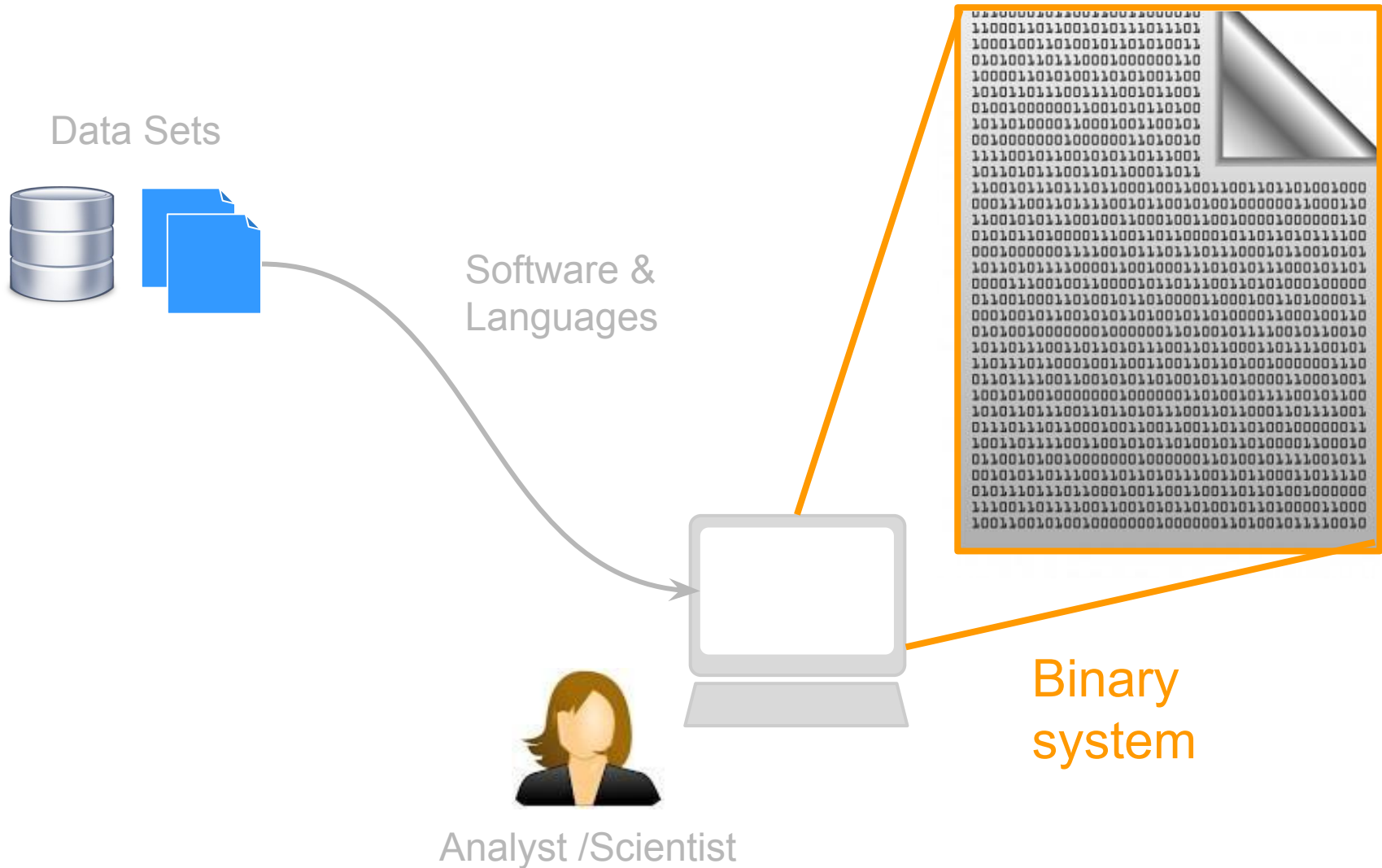
... Or setting up a data base

How do computers
treat data?





How computers treat Data



Data (for computers)

At the lowest level, computers treat all kinds of data in **binary** format:

0's and 1's

What about the
storage of **data**?

Data Technologies

Data Sets



Per Game

Share & more ▼

Glossary

Rk		Age	G	GS	MP	FG	FGA
1	Klay Thompson	26	78	78	34.0	8.3	17.6
2	Kevin Durant	28	62	62	33.4	8.9	16.5
3	Stephen Curry	28	79	79	33.4	8.5	18.3
4	Draymond Green	26	76	76	32.5	3.6	8.6
5	Andre Iguodala	33	76	0	26.3	2.9	5.5
6	Matt Barnes	36	20	5	20.5	1.9	4.5
7	Zaza Pachulia	32	70	70	18.1	2.3	4.4
8	Shaun Livingston	31	76	3	17.7	2.3	4.2
9	Patrick McCaw	21	71	20	15.1	1.5	3.5
10	Ian Clark	25	77	0	14.8	2.7	5.6
11	David West	36	68	0	12.6	2.0	3.7
12	JaVale McGee	29	77	10	9.6	2.7	4.1
13	James Michael McAdoo	24	52	2	8.8	1.2	2.3
14	Damian Jones	21	10	0	8.5	0.8	1.6
15	Kevon Looney	20	53	4	8.4	1.1	2.0
16	Anderson Varejao	34	14	1	6.6	0.4	1.0
17	Briante Weber	24	7	0	6.6	0.7	2.0

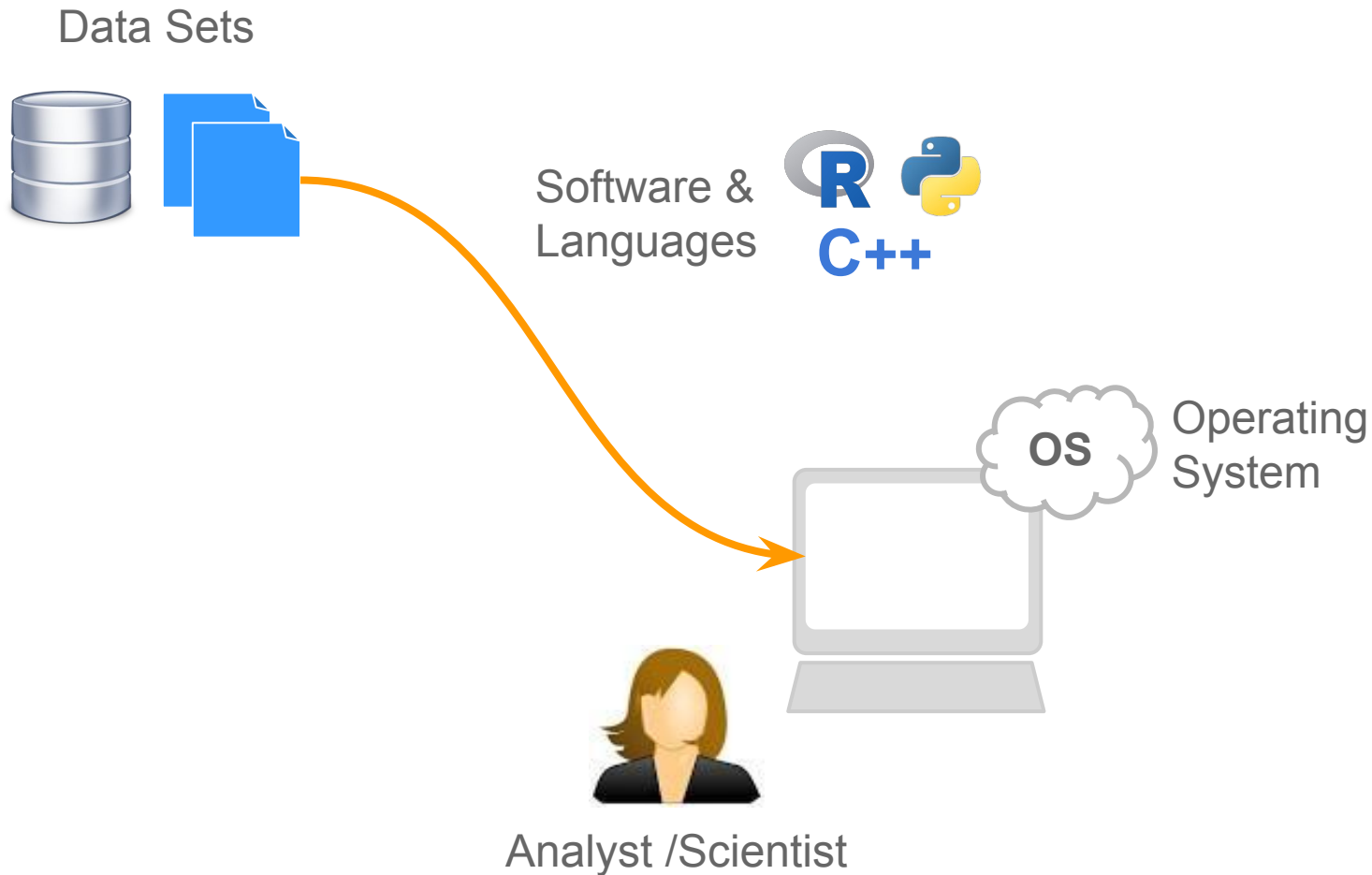
- File Format
- Encoding
- Metadata
- Location
- Size

Salaries

Rk		Salary
1	Kevin Durant	\$26,540,100
2	Klay Thompson	\$16,663,575
3	Draymond Green	\$15,330,435
4	Stephen Curry	\$12,112,359
5	Andre Iguodala	\$11,131,368
6	Shaun Livingston	\$5,782,450
7	Zaza Pachulia	\$2,898,000
8	David West	\$1,551,659
9	Anderson Varejao	\$1,551,659
10	JaVale McGee	\$1,403,611
11	Kevon Looney	\$1,182,840
12	Damian Jones	\$1,171,560
13	Ian Clark	\$1,015,696
14	James Michael McAdoo	\$980,431
15	Jason Thompson	\$945,126
16	Patrick McCaw	\$543,471
17	Matt Barnes	\$383,351
18	Elliot Williams	\$250,000

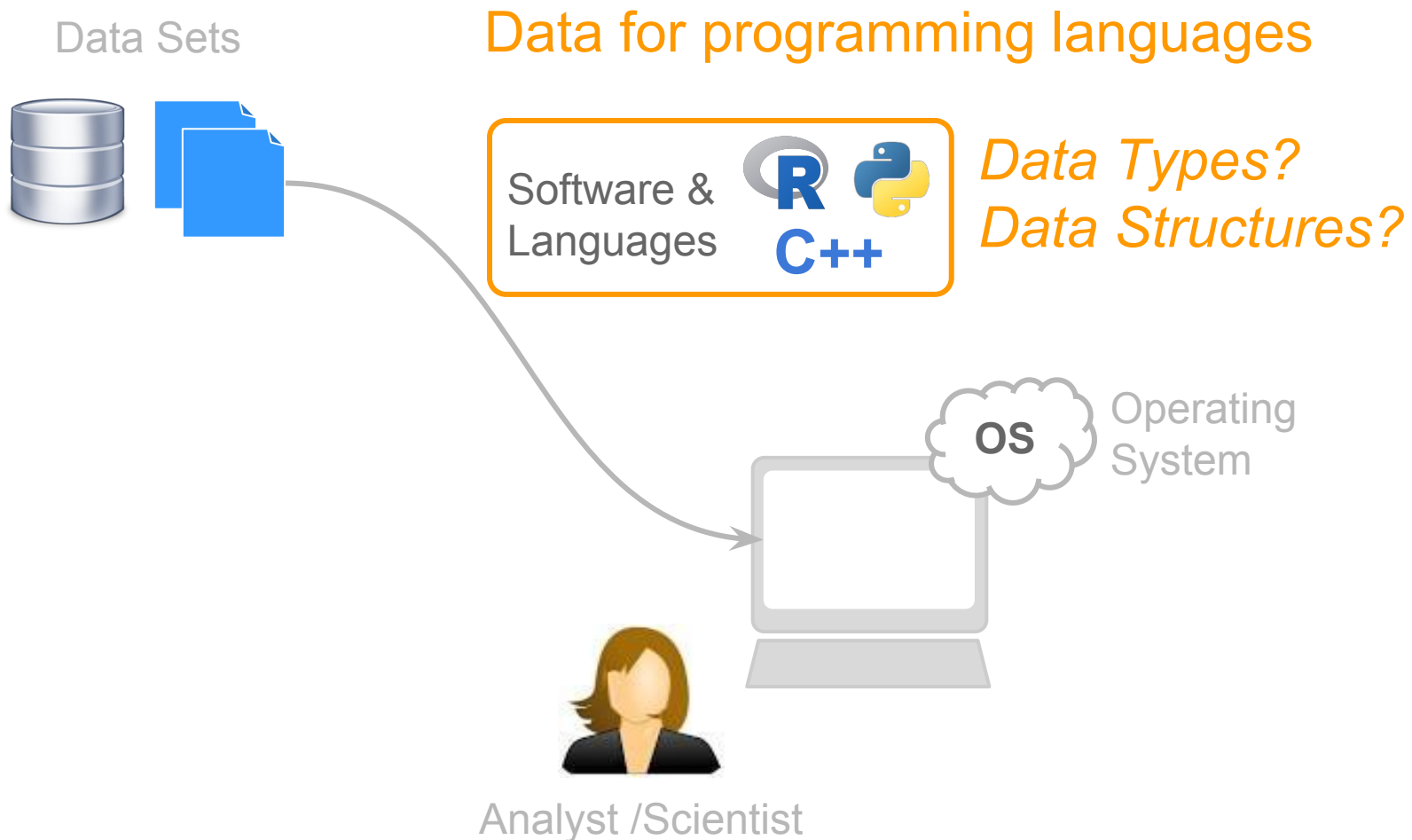
Importing & getting
access to **data**?

Importing/Accessing Data



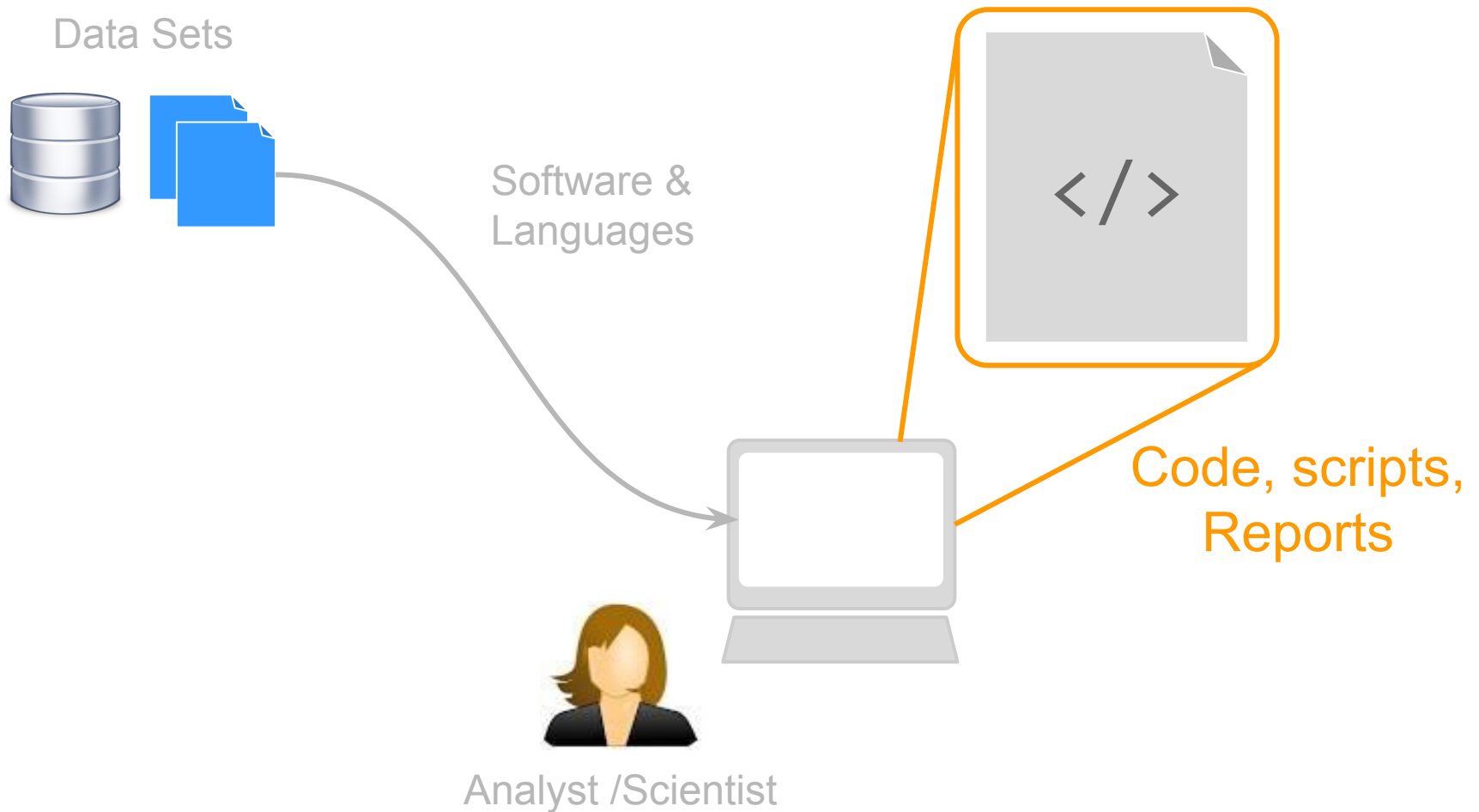
How do programming
languages handle data?

How Program/Languages handle Data



And what about all the
instructions, analysis,
reports?

Instructions, Scripts & Programs



Last but not least ...

Displayed/Reported Data

player	number	team	position	height	weight	birth_date	country
Andre Iguodala	9	GSW	SF	6-Jun	215	28-Jan-84	us
Damian Jones	15	GSW	C	Jul-00	245	30-Jun-95	us
David West	3	GSW	C	9-Jun	250	29-Aug-80	us
Draymond Green	23	GSW	PF	7-Jun	230	4-Mar-90	us
Ian Clark	21	GSW	SG	3-Jun	175	7-Mar-91	us
James Michael McAdoo	20	GSW	PF	9-Jun	230	4-Jan-93	us
JaVale McGee	1	GSW	C	Jul-00	270	19-Jan-88	us
Kevin Durant	35	GSW	SF	9-Jun	240	29-Sep-88	us
Kevon Looney	5	GSW	C	9-Jun	220	6-Feb-96	us
Klay Thompson	11	GSW	SG	7-Jun	215	8-Feb-90	us
Matt Barnes	22	GSW	SF	7-Jun	226	9-Mar-80	us
Patrick McCaw	0	GSW	SG	7-Jun	185	25-Oct-95	us
Shaun Livingston	34	GSW	PG	7-Jun	192	11-Sep-85	us
Stephen Curry	30	GSW	PG	3-Jun	190	14-Mar-88	us
Zaza Pachulia	27	GSW	C	11-Jun	270	10-Feb-84	ge



The Data Computing Diagram (DCD)

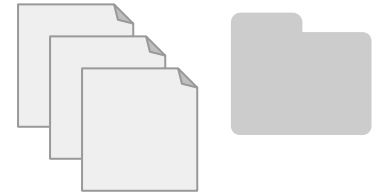
Data
Sets



Software &
Languages



Code, Scripts,
Programs



Computers



Analyst /Scientist

