

# The 3 Views of Data

---

Stat 133 with 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 looked at the DAC (very informally) using  
Long Jump World Record Progression

# The Data Computing Diagram (DCD)

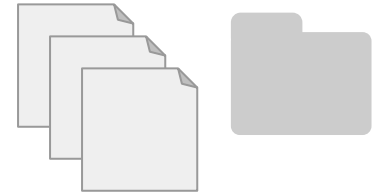
Data  
Sets



Software &  
Languages



Code, Scripts,  
Programs



Computers



Analyst /Scientist

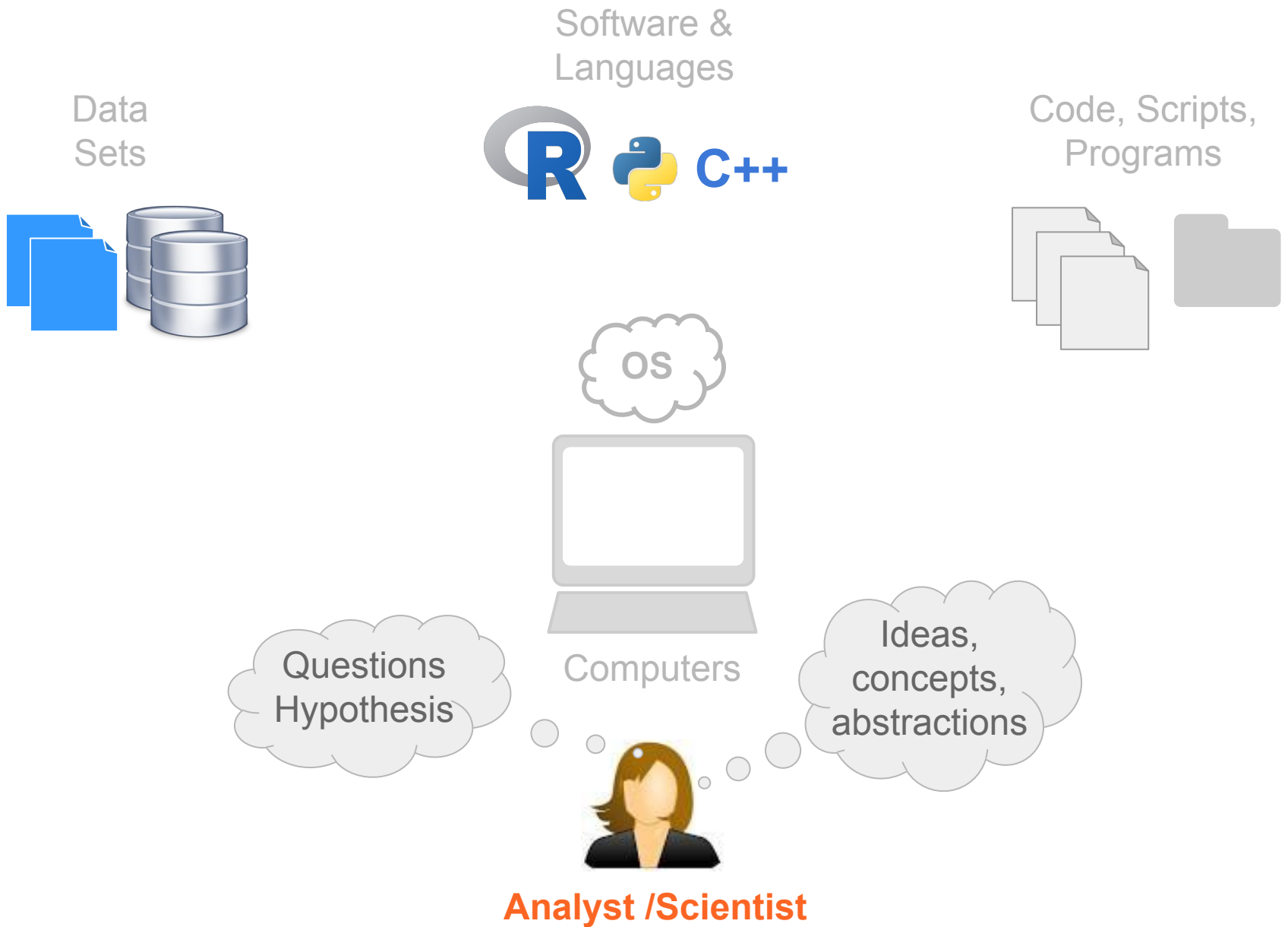
# Research Question



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 Perspectives



Quantitative  
variables



Analyst /Scientist

# Statistical Perspectives

```
r = cor(Salary, Points)
```

Summaries



Analyst /Scientist

# Statistical Perspectives

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

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

Theoretical  
Model



Analyst /Scientist

# Statistical Perspectives

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

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

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](http://basketball-reference.com)

# Roster

Share & more ▼

[Glossary](#)

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

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

# Roster

Share & more ▼

Glossary

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



# Per Game

Share & more ▼

Glossary

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

## Salaries

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

# Data Collection

Do Data need to be collected?

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



In this course we will assume that ...

Data is already in digital form

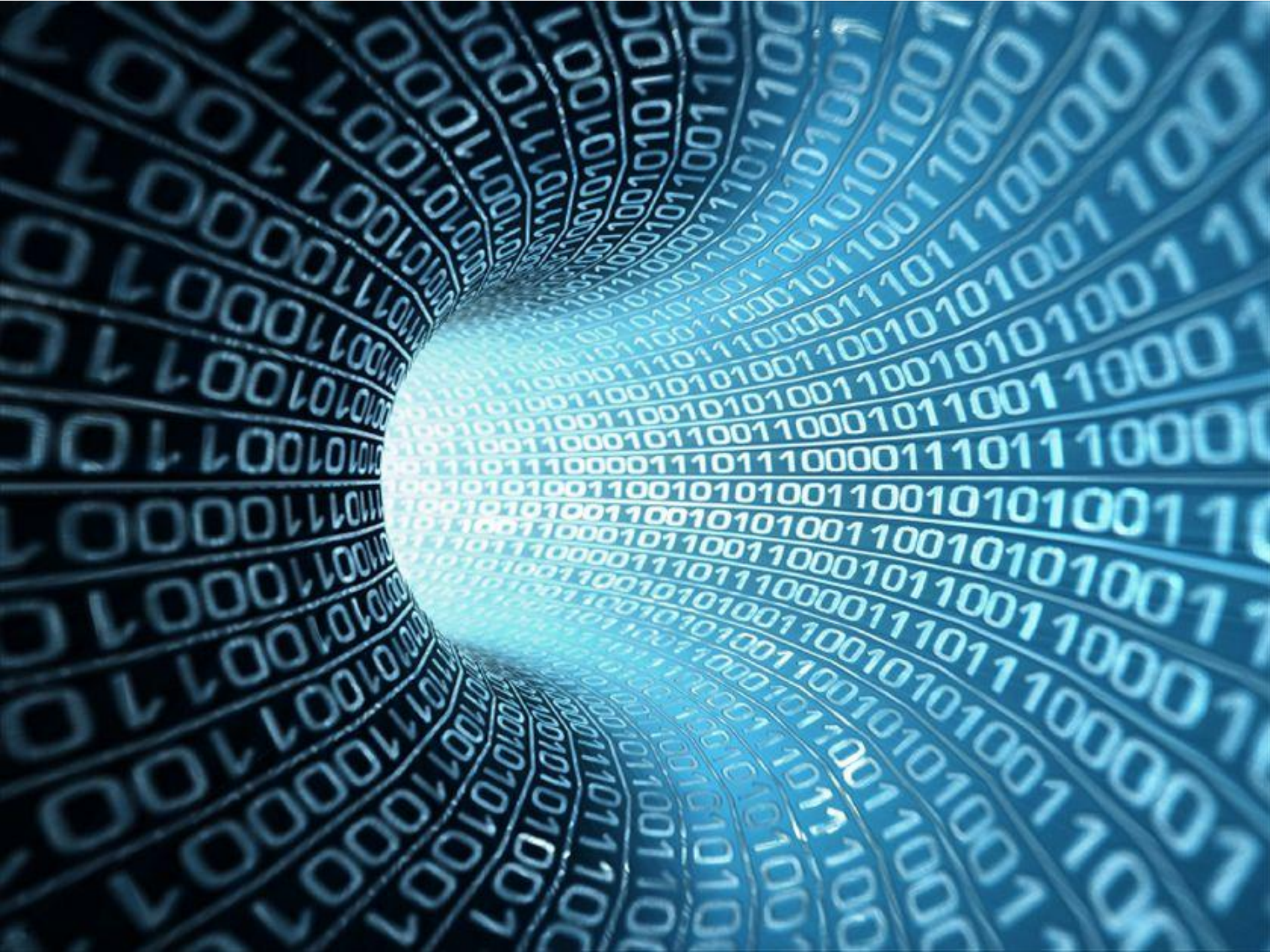
It has already been collected

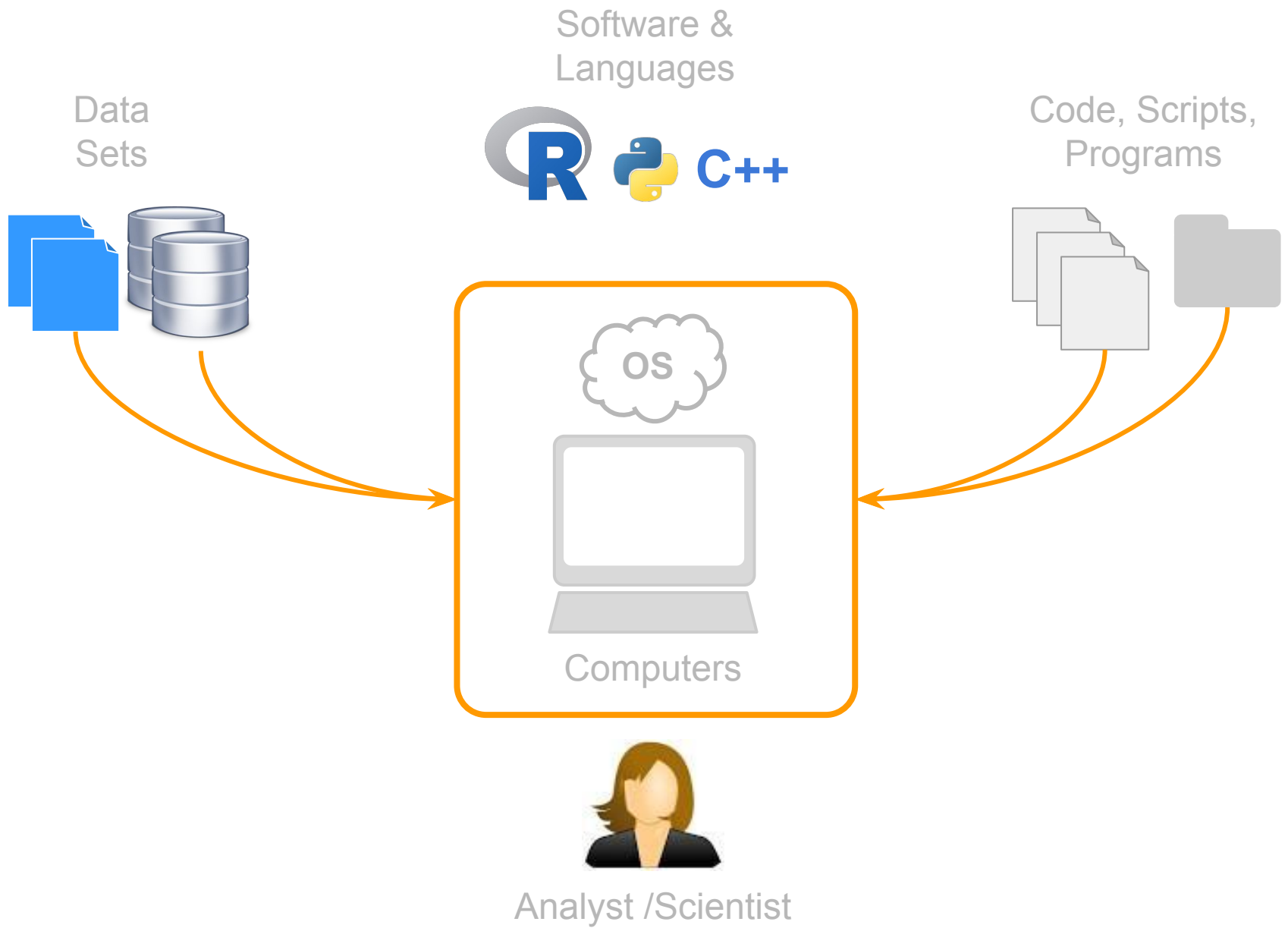
It is already stored in some files/directories

No worries about transcribing data

... Or setting up a data base

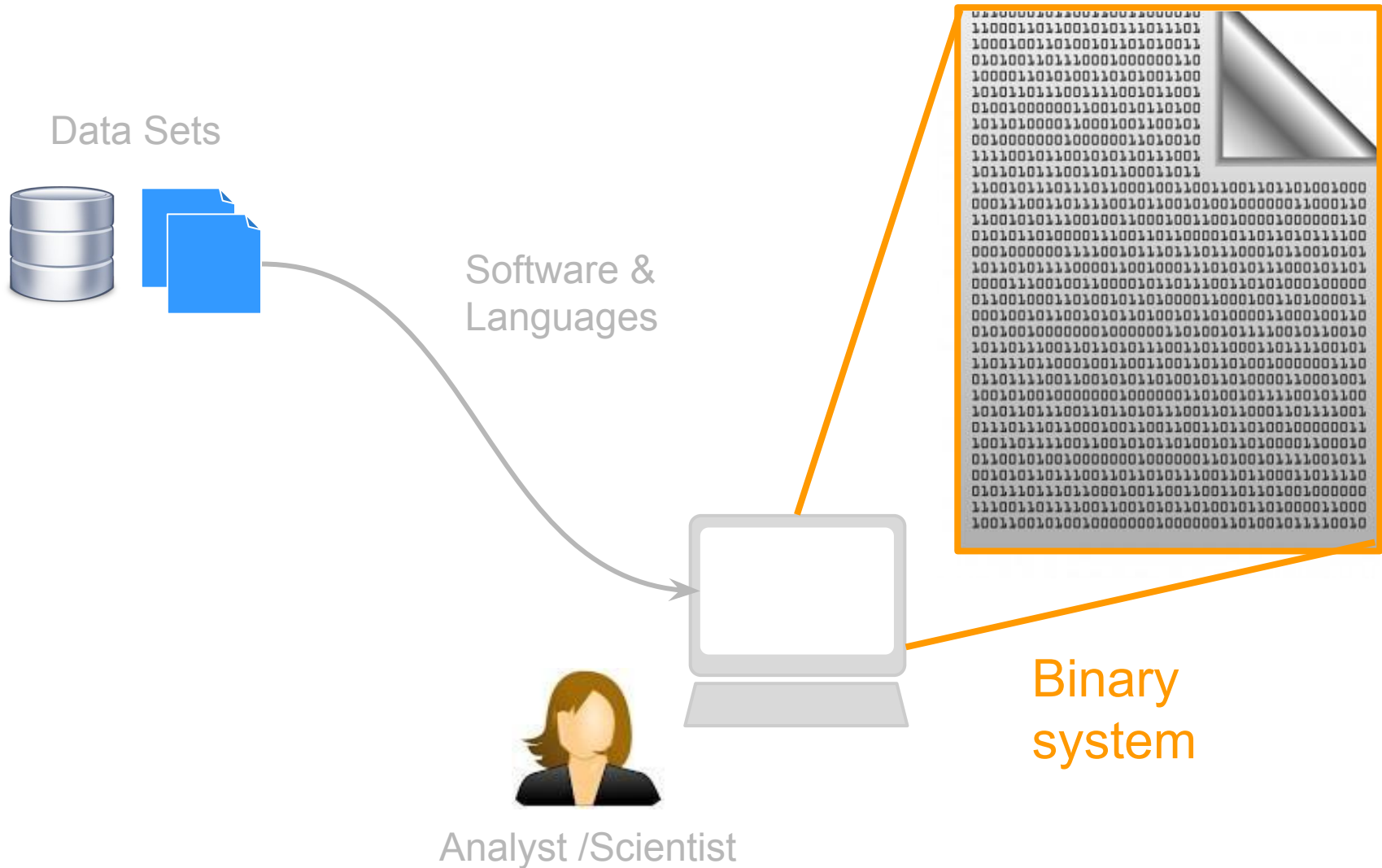
How do computers  
treat data?







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

- File Format
- Encoding
- Metadata
- Location
- Size

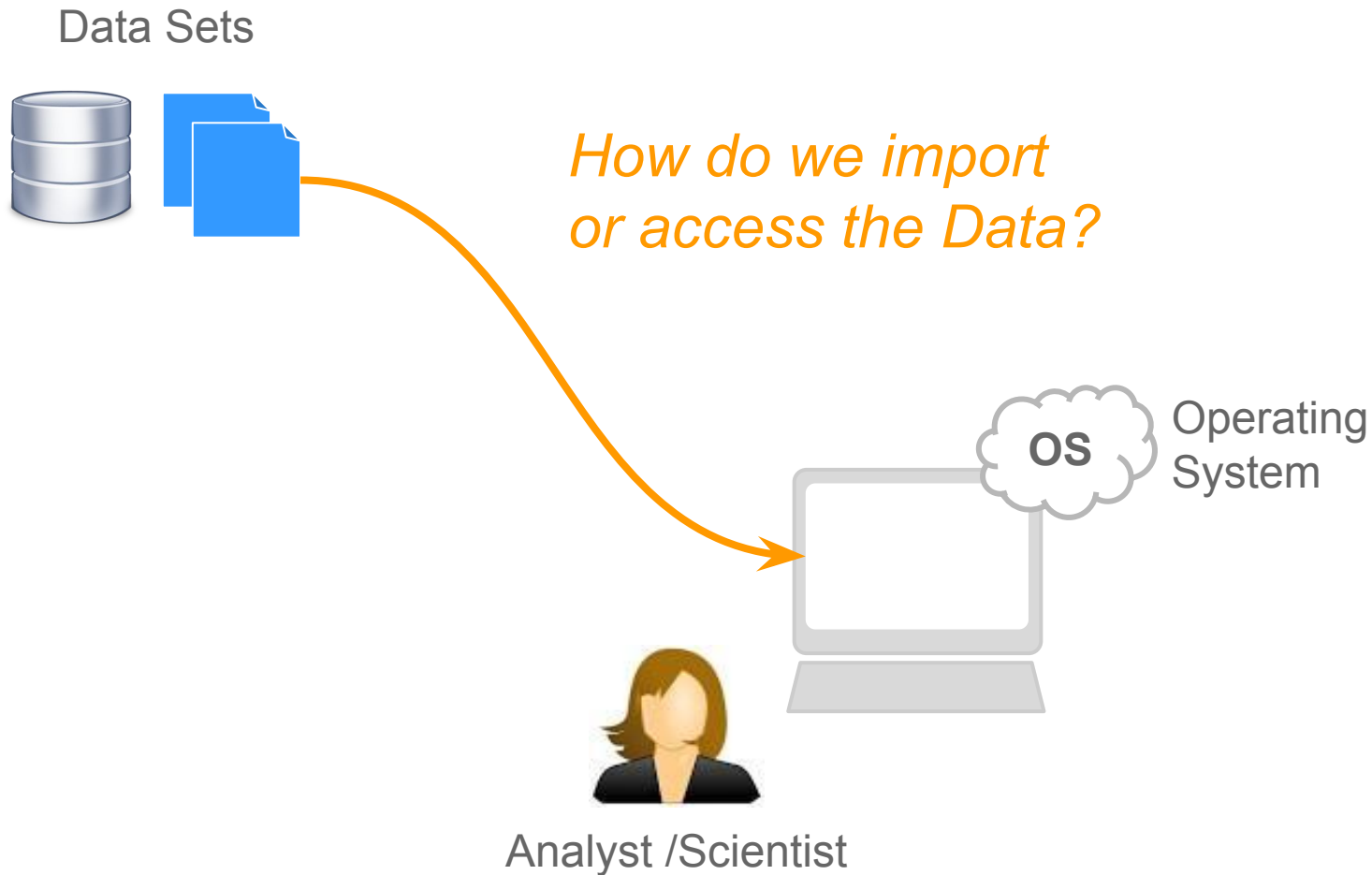
### Salaries

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



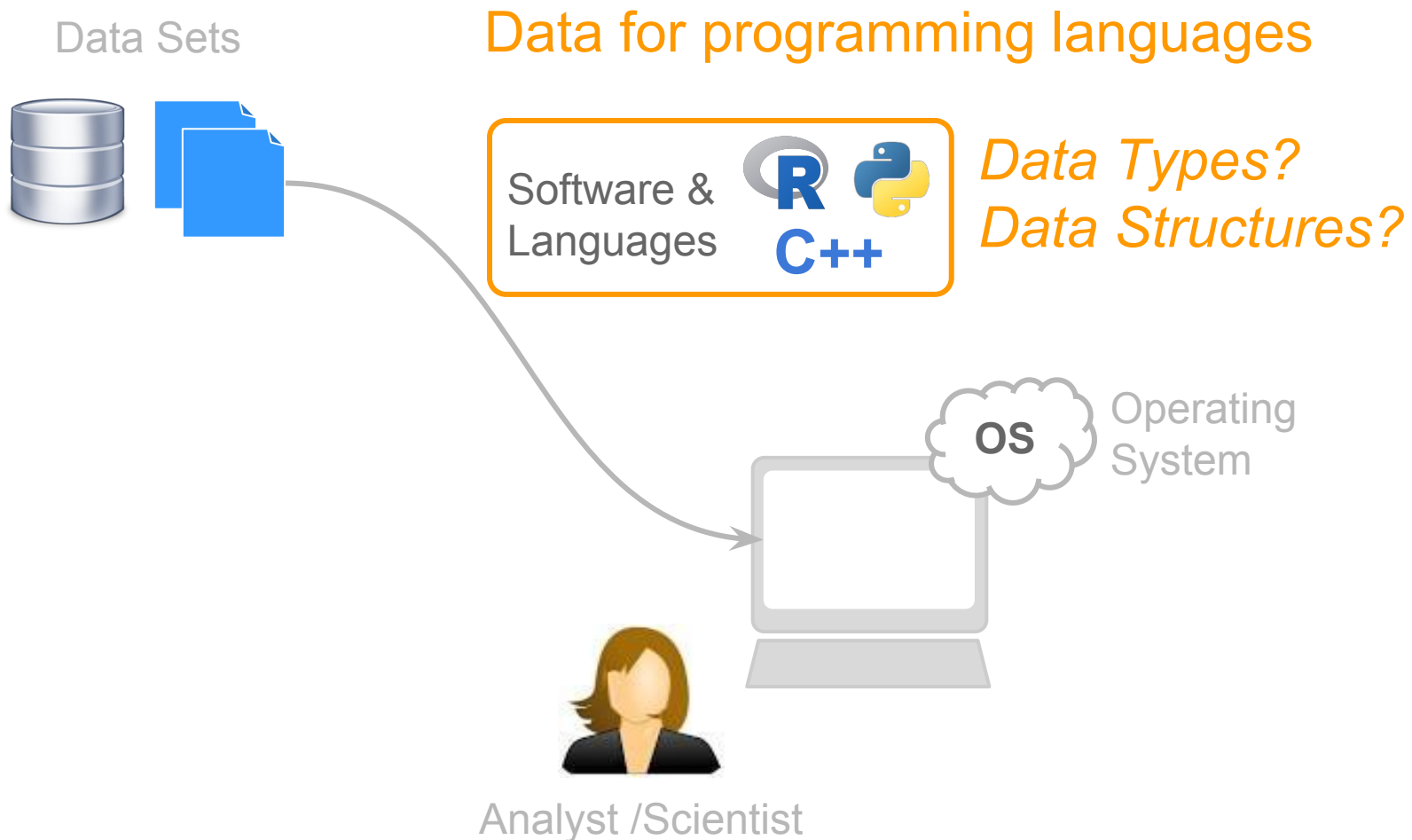
Importing & getting  
**access** to **data**?

# Importing/Accessing Data



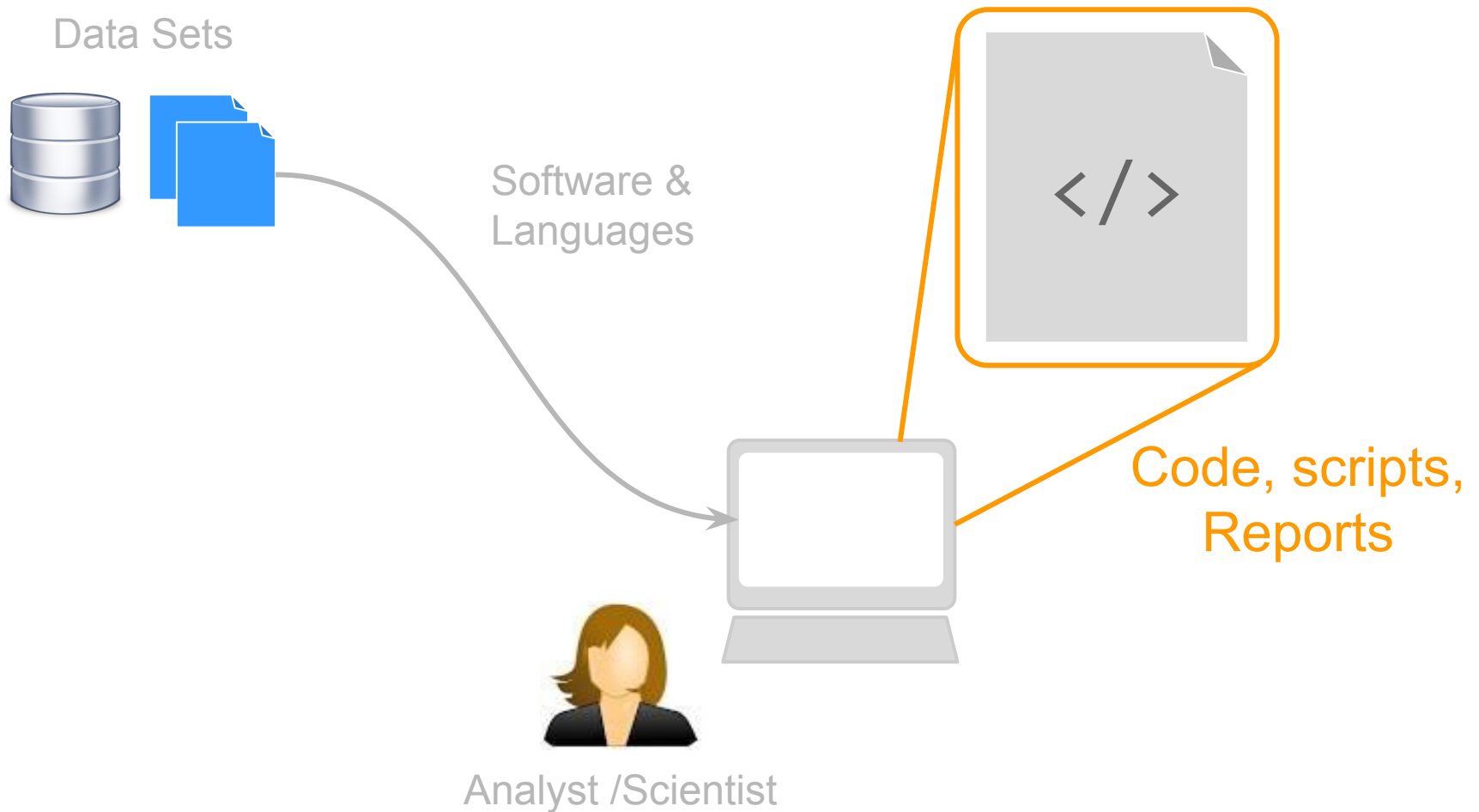
How do programming  
languages handle data?

# 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)

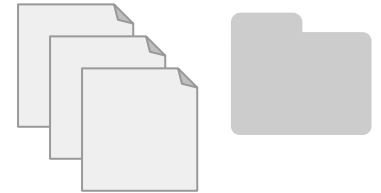
Data  
Sets



Software &  
Languages



Code, Scripts,  
Programs



Computers



Analyst /Scientist

