

Sum of Nothing

R, Python pandas, and the different ways to do math with missing data

Christine Zhang (@christinezhang)

[STATISTICS](#) / [DATA SCIENCE](#)

Data scientists mostly just do arithmetic

Feb 18, 2016

Noah Lorang, a data scientist at Basecamp, explains the key for most companies isn't finding a way to use the most advanced methods. Instead, it's about asking the right questions.

**Sometimes I'm called a "data scientist."
Mostly, I just do arithmetic, and I'm ok with that.**

— Noah Lorang

df

x	y	z
1	1	NA
2	2	NA
3	NA	NA

df

x	y	z
1	1	NA
2	2	NA
3	NA	NA



```
> sum(df$x)
```

```
[1] 6
```

```
> sum(df$y)
```

```
[1] NA
```

```
> sum(df$z)
```

```
[1] NA
```


df

x	y	z
1	1	NA
2	2	NA
3	NA	NA



```
> sum(df$x)
```

```
[1] 6
```

```
> sum(df$y, na.rm = TRUE)
```

```
[1] 3
```

```
> sum(df$z, na.rm = TRUE)
```

```
[1] 0
```

df

x	y	z
1	1	NaN
2	2	NaN
3	NaN	NaN

Pandas



```
>>> df['x'].sum()
6
>>> df['y'].sum()
3.0
>>> df['z'].sum()
0.0
```


df

x	y	z
1	1	NaN
2	2	NaN
3	NaN	NaN



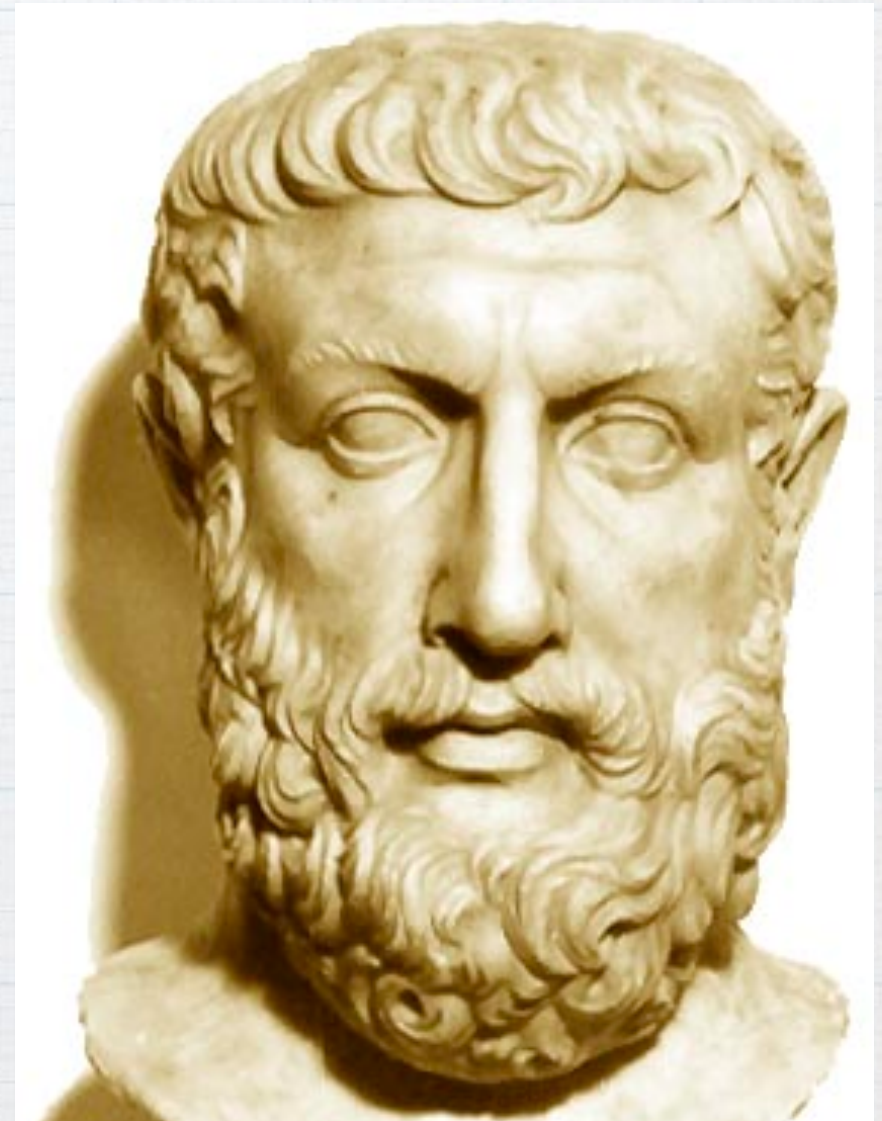
Pandas

```
>>> df['x'].sum(min_count = 1)
6
>>> df['y'].sum(min_count = 1)
3.0
>>> df['z'].sum(min_count = 1)
```

nan

**Something cannot come
from nothing**

**— Parmenides
(b. 515 BC)**



df

z
NA
NA
NA

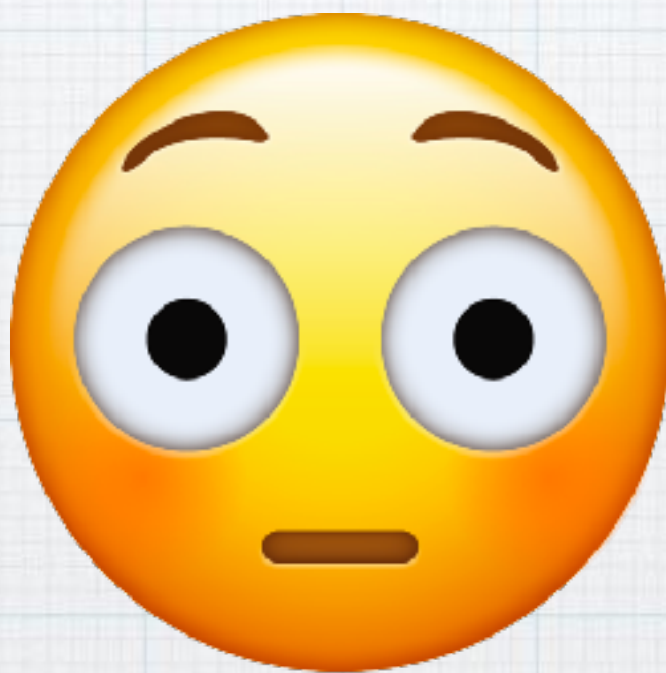
na.rm = TRUE
(or equivalent)
happens in
Python pandas
by default

```
> sum(df$z, na.rm = TRUE)  
[1] 0
```

```
>>> df['z'].sum()  
0.0
```



MATH



sum

From [base v3.5.0](#)
by [R-core R-core@R-project.org](mailto:R-core@R-project.org)

Sum Of Vector Elements

`sum` returns the sum of all the values present in its arguments.

Keywords [arith](#)

Usage

```
sum(..., na.rm = FALSE)
```



NB: the sum of an empty set is zero, by definition.

<https://www.rdocumentation.org/packages/base/versions/3.5.0/topics/sum>

Empty sum

In **mathematics**, an **empty sum**, or **nullary sum**, is a **summation** where the number of terms is zero. By convention,^[1] the value of any empty sum of numbers is the **additive identity**, **zero**.

Empty sum

In **mathematics**, an **empty sum**, or **nullary sum**, is a **summation** where the number of terms is zero. By convention,^[1] the value of any empty sum of numbers is the **additive identity**, **zero**.

df

z
NA
NA
NA

```
> sum(df$z, na.rm = TRUE)
[1] 0
```

Empty sum

In **mathematics**, an **empty sum**, or **nullary sum**, is a **summation** where the number of terms is zero. By convention,^[1] the value of any empty sum of numbers is the **additive identity, zero**.

df

z
NA
NA
NA

```
> sum()  
[1] 0
```

```
> sum(df$z, na.rm = TRUE)  
[1] 0
```


df

x	y	z
1	1	NA
2	2	NA
3	NA	NA



```
> prod(df$x, na.rm = TRUE)
[1] 6
> prod(df$y, na.rm = TRUE)
[1] 2
> prod(df$z, na.rm = TRUE)
```

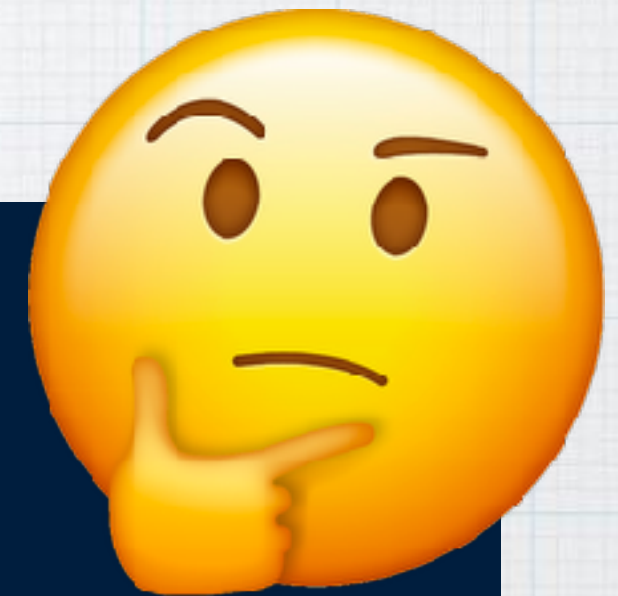
?

df

x	y	z
1	1	NA
2	2	NA
3	NA	NA



```
> prod(df$x, na.rm = TRUE)
[1] 6
> prod(df$y, na.rm = TRUE)
[1] 2
> prod(df$z, na.rm = TRUE)
[1] 1
```



df

x	y	z
1	1	NA
2	2	NA
3	NA	NA

Pandas



```
>>> df['x'].prod()
6
>>> df['y'].prod()
2.0
>>> df['z'].prod()
1.0
```



prod

From [base v3.5.0](#)
by [R-core R-core@R-project.org](#)

Product Of Vector Elements

`prod` returns the product of all the values present in its arguments.

Keywords [arith](#)

Usage

```
prod(..., na.rm = FALSE)
```



NB: the product of an empty set is one, by definition.

<https://www.rdocumentation.org/packages/base/versions/3.5.0/topics/prod>

Empty product

In **mathematics**, an **empty product**, or **nullary product**, is the result of **multiplying** no factors. It is by convention equal to the multiplicative **identity 1** (assuming there is an identity for the multiplication operation in question), just as the **empty sum**—the result of **adding** no numbers—is by convention **zero**, or the additive identity.^{[1][2][3][4]}

Empty product

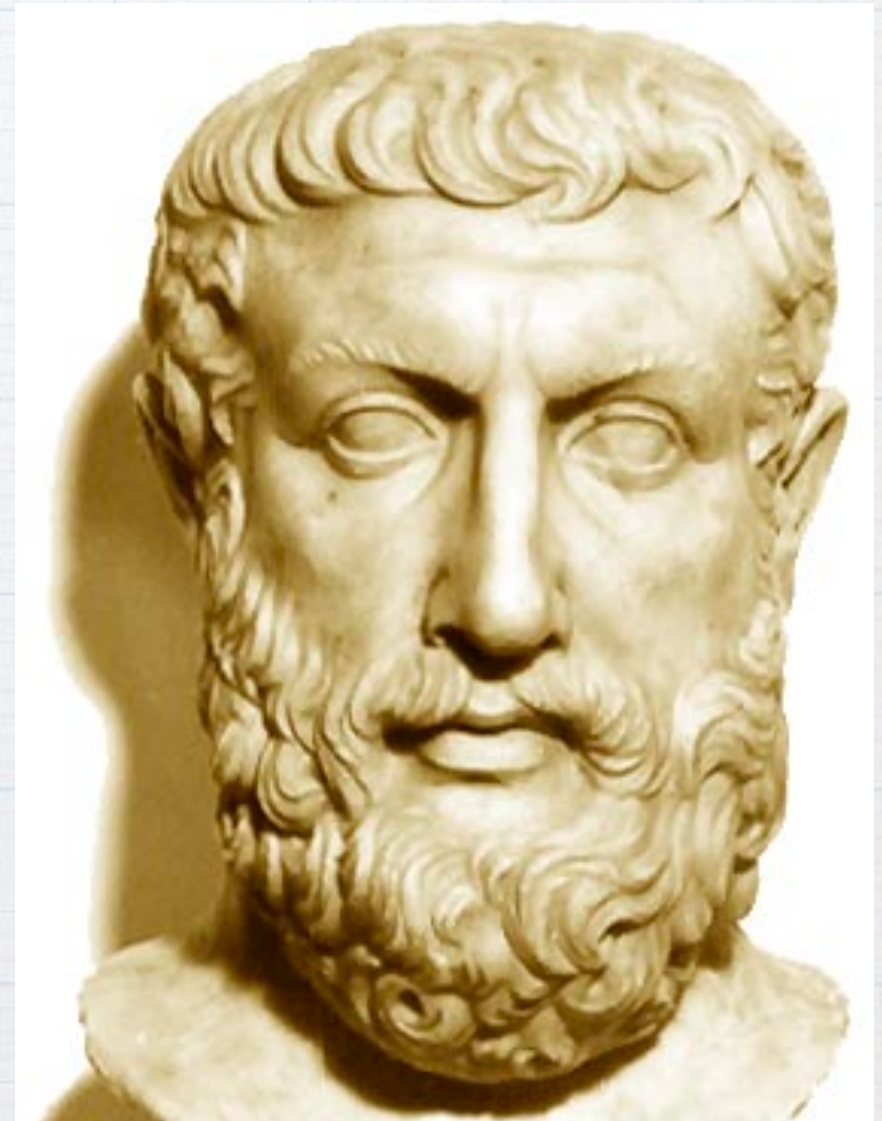
$$\begin{aligned} e &= e^1 \\ &= e^{(1 + 0)} \\ &= e^1 \times e^0 \end{aligned}$$



the “empty product”

**Something cannot come
from nothing**

**— Parmenides
(b. 515 BC)**

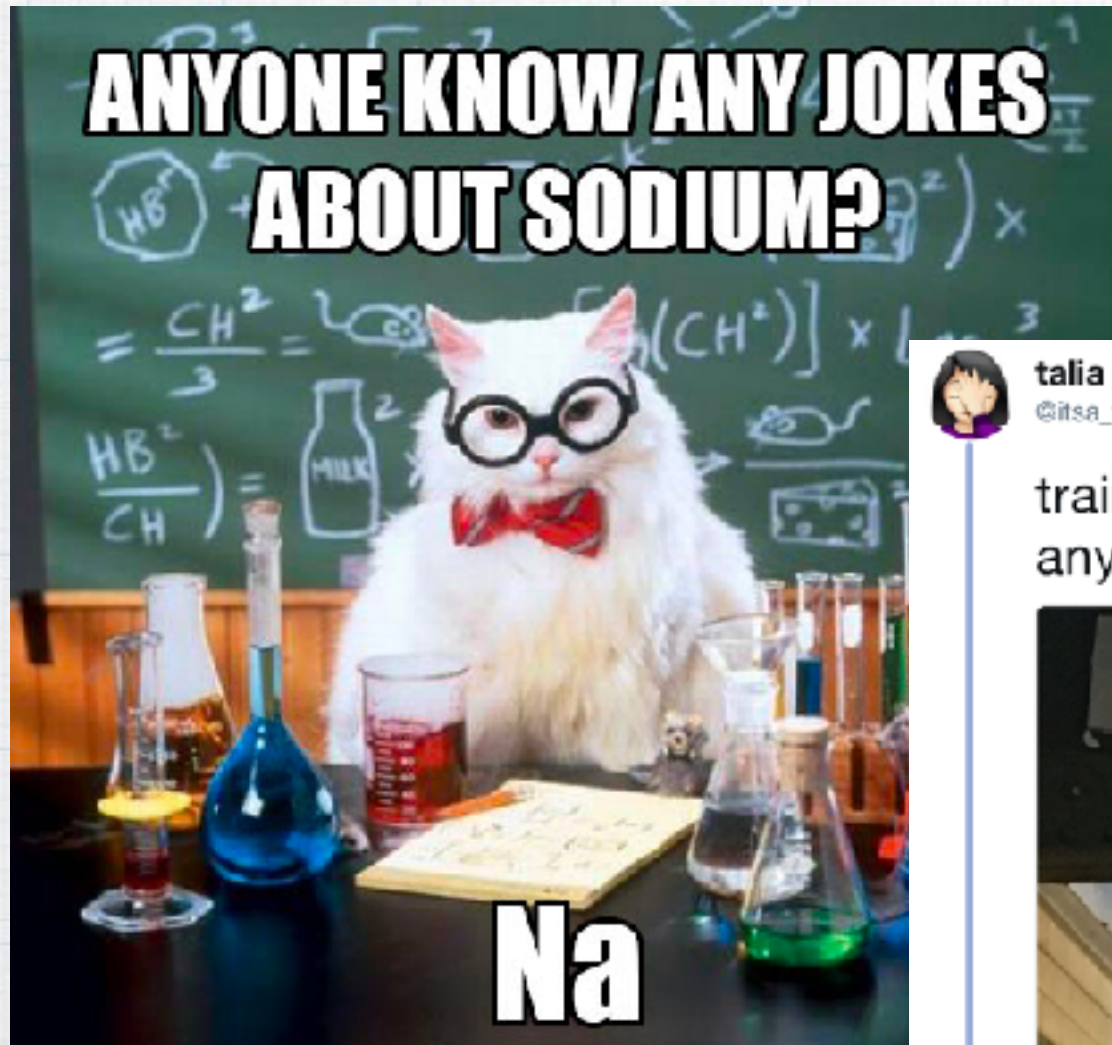


**At first there was nothing ...
then nothing turned itself
inside-out and became
something**

**—Sun Ra (b. 1934)
(and Yo La Tengo)**



Thank you



@christinezhang

ychristinezhang
at gmail dot com

