Software Choices:

Lab 1

Agenda

- Why write code?
- Choosing a programming language
- What is “data”?
- You are not alone (how to find help)

- Overview of the RStudio environment
- R programming basics and best practices

Why programming? Why not use Excel?

I3 | fx | 12, 8

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	EventID	EventType	StartDateTim	EndDateTim	EnteredOn	EventAgency	ParkingHeld	Borough	CommunityB	PolicePrecinc	Category	SubCategory	Country	ZipCode(s)	
2	446040	Shooting Per	##########	##########	##########	Mayor's Offi	THOMPSON	Manhattan	2	1	Television	Cable-episod	United State	10012	
3	446168	Shooting Per	##########	##########	##########	Mayor's Offi	MARBLE HILI	Manhattan	12, 8	34, 50	Film	Feature	United State	10034, 10463	
4	186438	Shooting Per	##########	##########	##########	Mayor's Offi	LAUREL HILL	Queens	2, 5	104, 108	Television	Episodic seri	United State	11378	
5	445255	Shooting Per	##########	##########	##########	Mayor's Offi	JORALEMON	Brooklyn	2	84	Still Photogra	Not Applicab	United State	11201	
6	128794	Theater Loac	##########	##########	##########	Mayor's Offi	WEST 31 ST	Manhattan	4, 5	14	Theater	Theater	United State	10001, 10121	
7	43547	Shooting Per	##########	##########	##########	Mayor's Offi	EAGLE STREI	Brooklyn	1, 2	108, 94	Television	Episodic seri	United State	11101, 11222	
8	66846	Shooting Per	##########	##########	##########	Mayor's Offi	8 AVENUE be	Brooklyn	6	78	Film	Feature	United State	11217	
9	104342	Shooting Per	##########	##########	##########	Mayor's Offi	WEST 44 ST	Manhattan	5	14	Television	News	United State	10036	
10	244863	Shooting Per	##########	##########	##########	Mayor's Offi	BRONXDALE	Bronx	11	49	Television	Cable-episod	United State	10462	
11	446379	Shooting Per	##########	##########	##########	Mayor's Offi	JANE STREET	Manhattan	2	6	WEB	Not Applicab	United State	10014	
12	446359	Shooting Per	##########	##########	##########	Mayor's Offi	WEST 48 ST	Manhattan	5	18	Television	News	United State	10036, 10105	
13	203743	Shooting Per	##########	##########	##########	Mayor's Offi	43 AVENUE I	Queens	2	108, 6	Still Photogra	Not Applicab	United State	10014, 11101	
14	446069	Shooting Per	##########	##########	##########	Mayor's Offi	EAST 37 ST	Brooklyn	14, 17	67, 70	Commercial	Promo	United State	11203, 11218	
15	445165	Theater Loac	##########	##########	##########	Mayor's Offi	WEST 31 ST	Manhattan	4, 5	14	Theater	Theater	United State	10001, 10121	
16	82397	Shooting Per	1/7/13 7:00	##########	##########	Mayor's Offi	13 AVENUE I	Brooklyn	12	66	Television	Episodic seri	United State	11219	

… 64,000 more rows

“90% of data ever created was in the last 2 years”

Why programming? Why not use Excel?

- Most tasks are too labor-intensive to do by hand
- Humans are prone to making mistakes
- Programs are documents of your analysis pipeline

Enter programming/scripting languages

Python
R
MATLAB
Julia
Java
SQL
Scala
SAS
Lua
Perl
PHP
C
C#
C++
Ruby

Banana Milkshake

Ingredients:

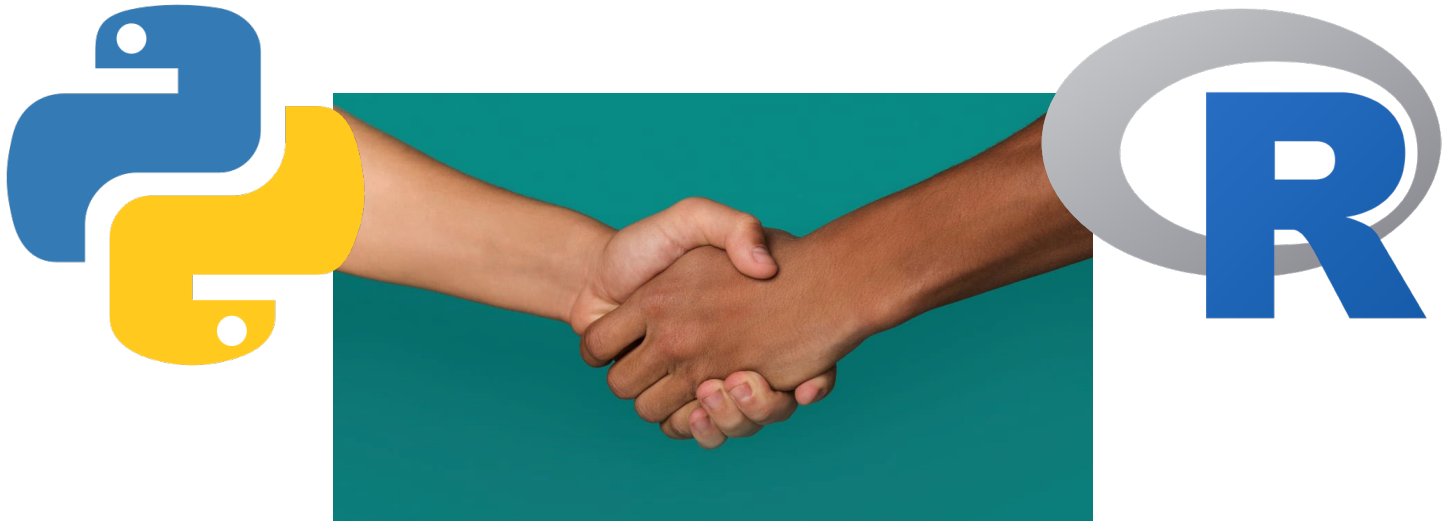
- 1 banana
- 250ml milk
- 3 ice cubes

Method:

1. Peel the banana and add to a blender
2. Add the milk and ice cubes to the blender
3. Turn on the blender for 30 seconds
4. Drink!

R versus Python

R and Python



- Both are useful, and worth your time exploring
- There's nothing one can do that the other can't
- Core programming concepts are same, main differences are:
  - Syntax
  - Community conventions
  - Package availability
- Learning one will ease learning others

R versus Python

```r
average = function(numbers)
{
  total = 0
  for (i in 1:length(numbers))
  {
    total = total + numbers[i]
  }
  mean = total / length(numbers)

  return(mean)
}
```

```
> average(c(1,2,3,4))
[1] 2.5
>
```

```python

def average(numbers):

    total = 0
    for i in range(len(numbers)):
        total = total + numbers[i]

    mean = total / len(numbers)

    return mean

```

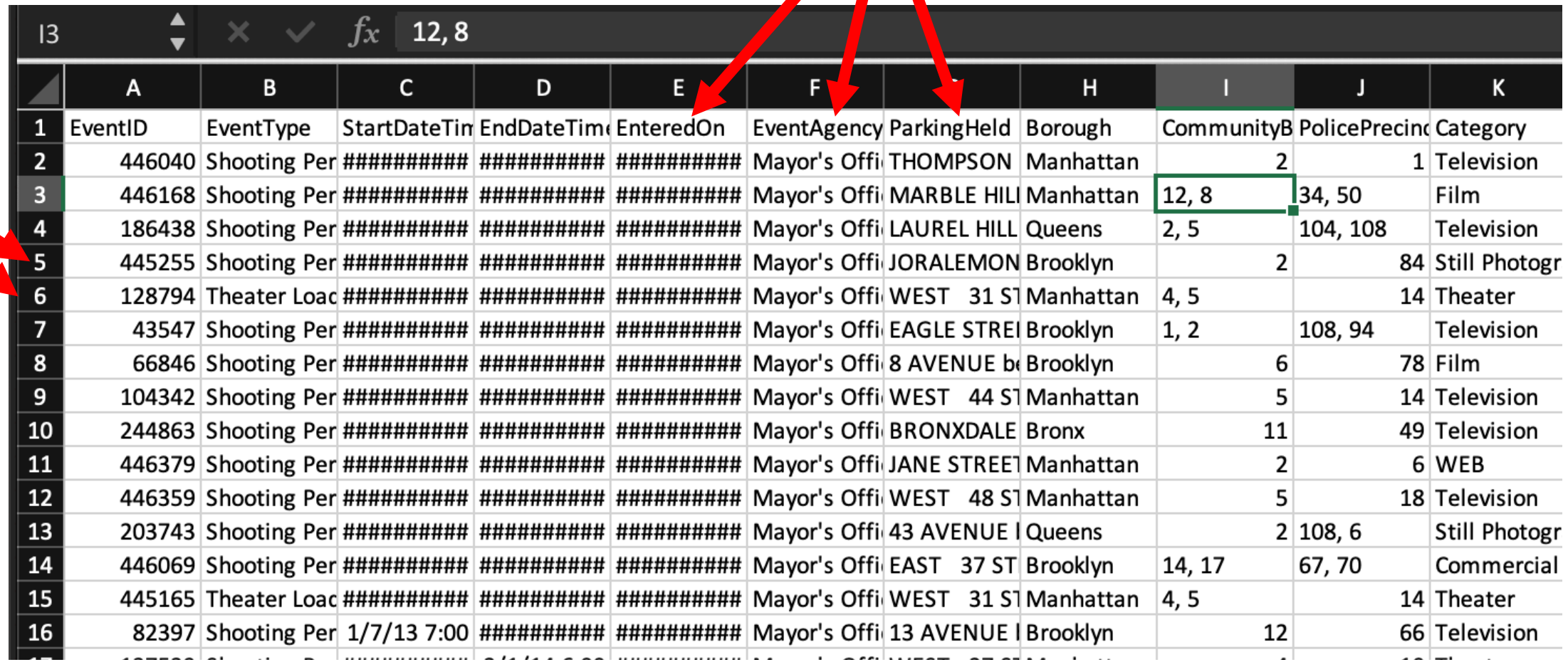
```
In [7]: average([1, 2, 3, 4])
Out[7]: 2.5
```

Where to find data

Data organization

columns (variables)

rows



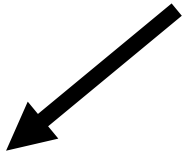
I3 | 12, 8

	A	B	C	D	E	F	G	H	I	J	K
1	EventID	EventType	StartDateTim	EndDateTime	EnteredOn	EventAgency	ParkingHeld	Borough	CommunityB	PolicePrecinc	Category
2	446040	Shooting Per	##########	##########	##########	Mayor's Offi	THOMPSON	Manhattan	2	1	Television
3	446168	Shooting Per	##########	##########	##########	Mayor's Offi	MARBLE HILL	Manhattan	12, 8	34, 50	Film
4	186438	Shooting Per	##########	##########	##########	Mayor's Offi	LAUREL HILL	Queens	2, 5	104, 108	Television
5	445255	Shooting Per	##########	##########	##########	Mayor's Offi	JORALEMON	Brooklyn	2	84	Still Photogr
6	128794	Theater Loac	##########	##########	##########	Mayor's Offi	WEST 31 ST	Manhattan	4, 5	14	Theater
7	43547	Shooting Per	##########	##########	##########	Mayor's Offi	EAGLE STREI	Brooklyn	1, 2	108, 94	Television
8	66846	Shooting Per	##########	##########	##########	Mayor's Offi	8 AVENUE be	Brooklyn	6	78	Film
9	104342	Shooting Per	##########	##########	##########	Mayor's Offi	WEST 44 ST	Manhattan	5	14	Television
10	244863	Shooting Per	##########	##########	##########	Mayor's Offi	BRONXDALE	Bronx	11	49	Television
11	446379	Shooting Per	##########	##########	##########	Mayor's Offi	JANE STREET	Manhattan	2	6	WEB
12	446359	Shooting Per	##########	##########	##########	Mayor's Offi	WEST 48 ST	Manhattan	5	18	Television
13	203743	Shooting Per	##########	##########	##########	Mayor's Offi	43 AVENUE I	Queens	2	108, 6	Still Photogr
14	446069	Shooting Per	##########	##########	##########	Mayor's Offi	EAST 37 ST	Brooklyn	14, 17	67, 70	Commercial
15	445165	Theater Loac	##########	##########	##########	Mayor's Offi	WEST 31 ST	Manhattan	4, 5	14	Theater
16	82397	Shooting Per	1/7/13 7:00	##########	##########	Mayor's Offi	13 AVENUE I	Brooklyn	12	66	Television

Common data formats

.txt

.bin



.csv

.tsv

.html

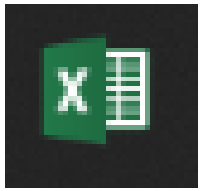
.json

.dat

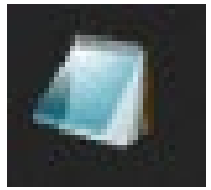
.jpg, .png, .tiff, .bmp

.wav

Common data formats

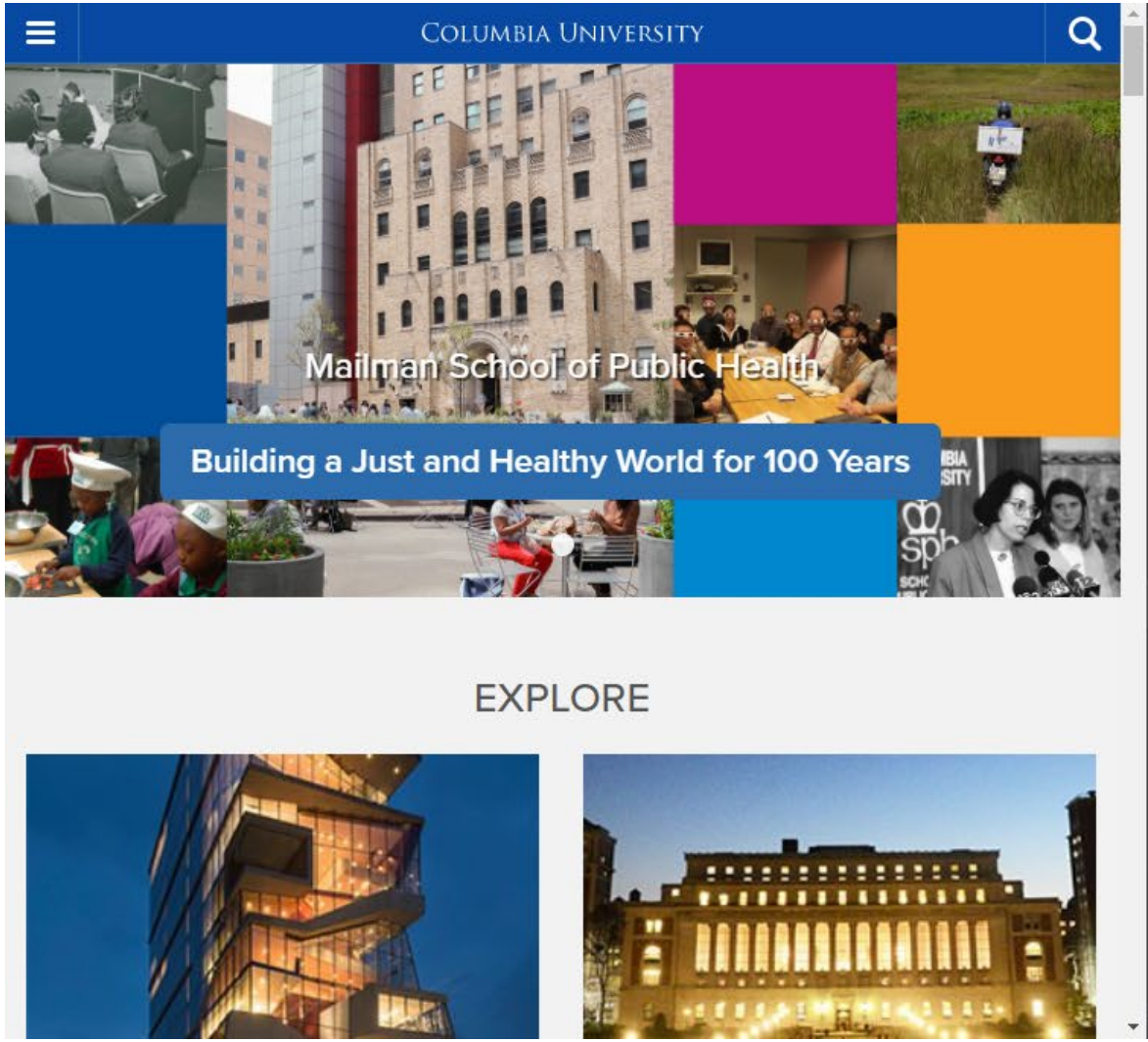


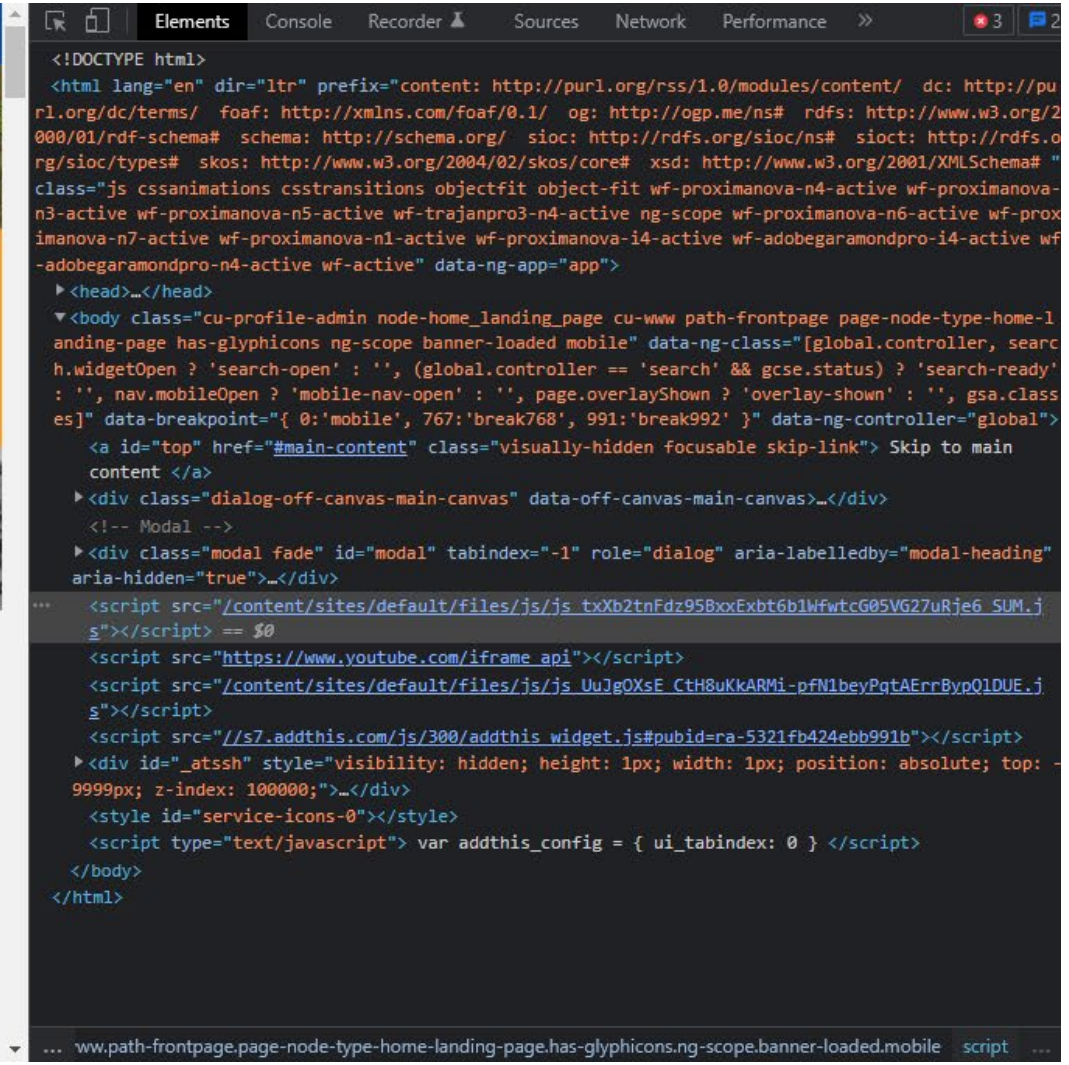
	A	B	C	D
1	songname	isi	species	duration
2	au2219bf_song01	NA	BF	0.065194
3	au2219bf_song01	0.118392	BF	0.066985
4	au2219bf_song01	0.0794118	BF	0.077439
5	au2219bf_song01	0.0782553	BF	0.07889
6	au2219bf_song01	0.0615884	BF	0.081634
7	au2219bf_song01	0.0093652	BF	0.09465
8	au2219bf_song01	0.0371208	BF	0.10719
9	au2219bf_song01	0.0120637	BF	0.056554
10	au2219bf_song01	0.0159413	BF	0.057892
11	au2219bf_song01	0.0144447	BF	0.064196



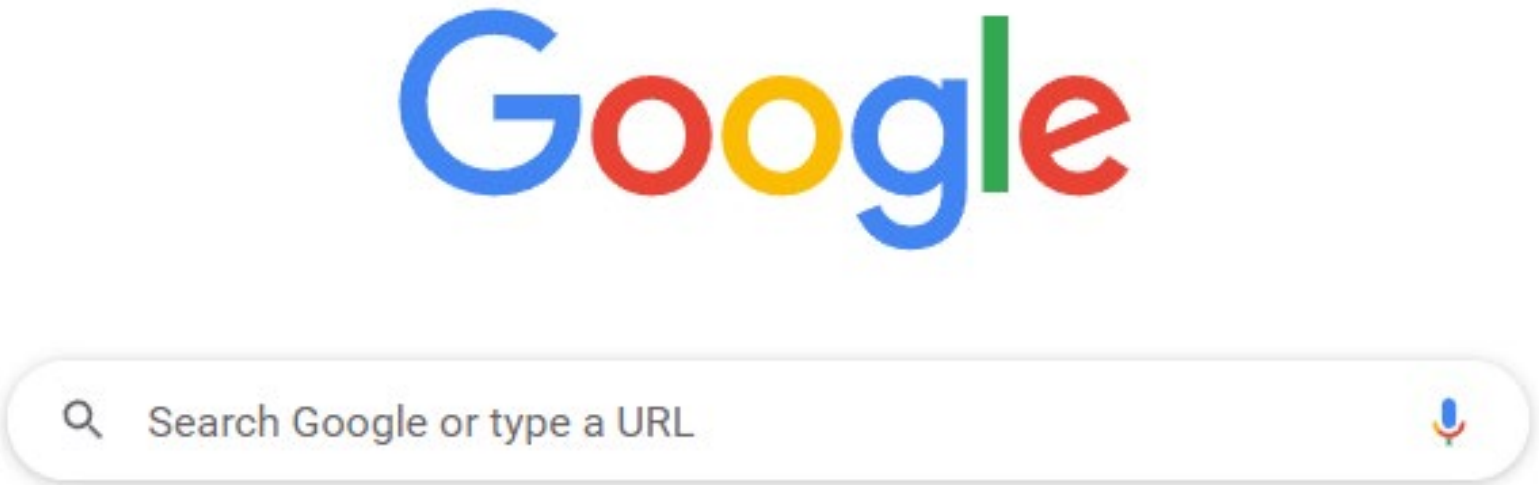
```
songname,isi,species,duration
au2219bf_song01,NA,BF,0.065193855
au2219bf_song01,0.118392041,BF,0.066985269
au2219bf_song01,0.079411785,BF,0.077438962
au2219bf_song01,0.078255302,BF,0.078890234
au2219bf_song01,0.061588352,BF,0.081634045
au2219bf_song01,0.009365239,BF,0.09465014
au2219bf_song01,0.037120814,BF,0.107190036
au2219bf_song01,0.012063697,BF,0.056554252
au2219bf_song01,0.015941315,BF,0.057892144
au2219bf_song01,0.014444691,BF,0.064196106
```

Common data formats

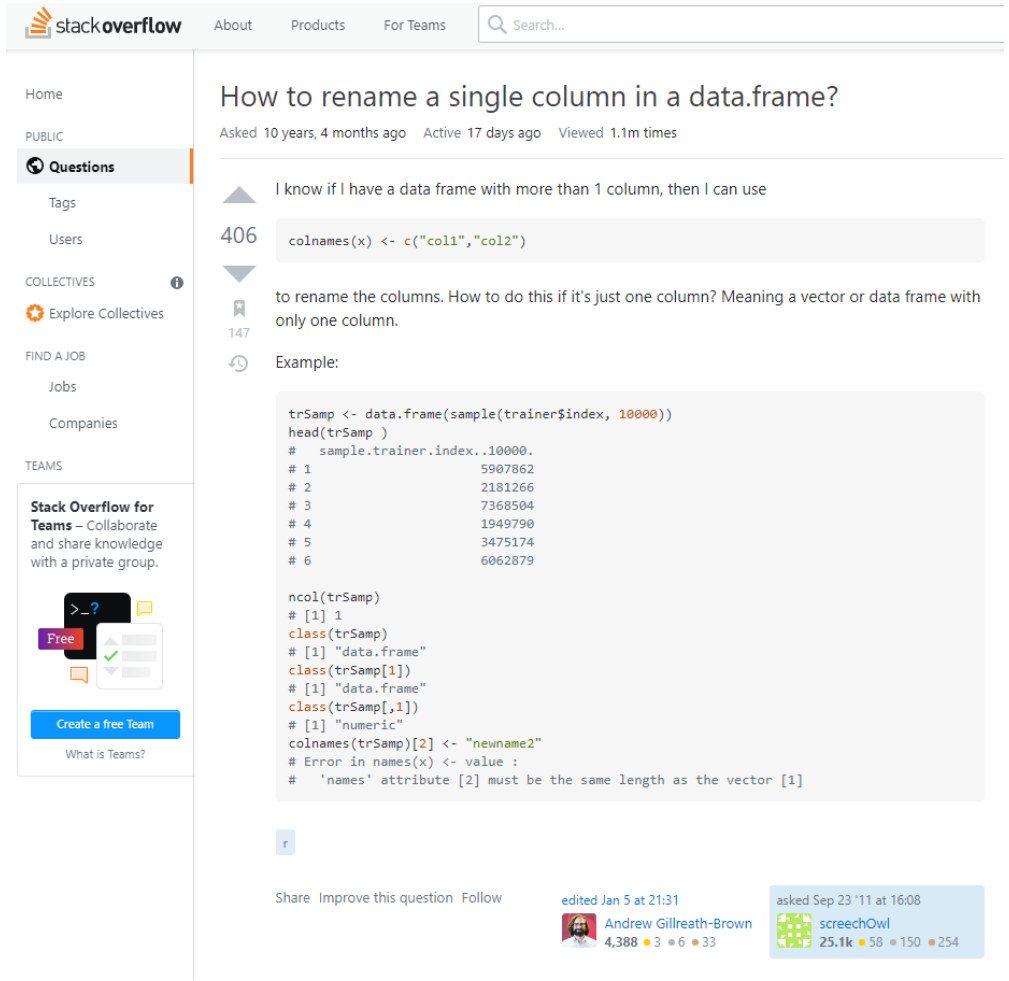


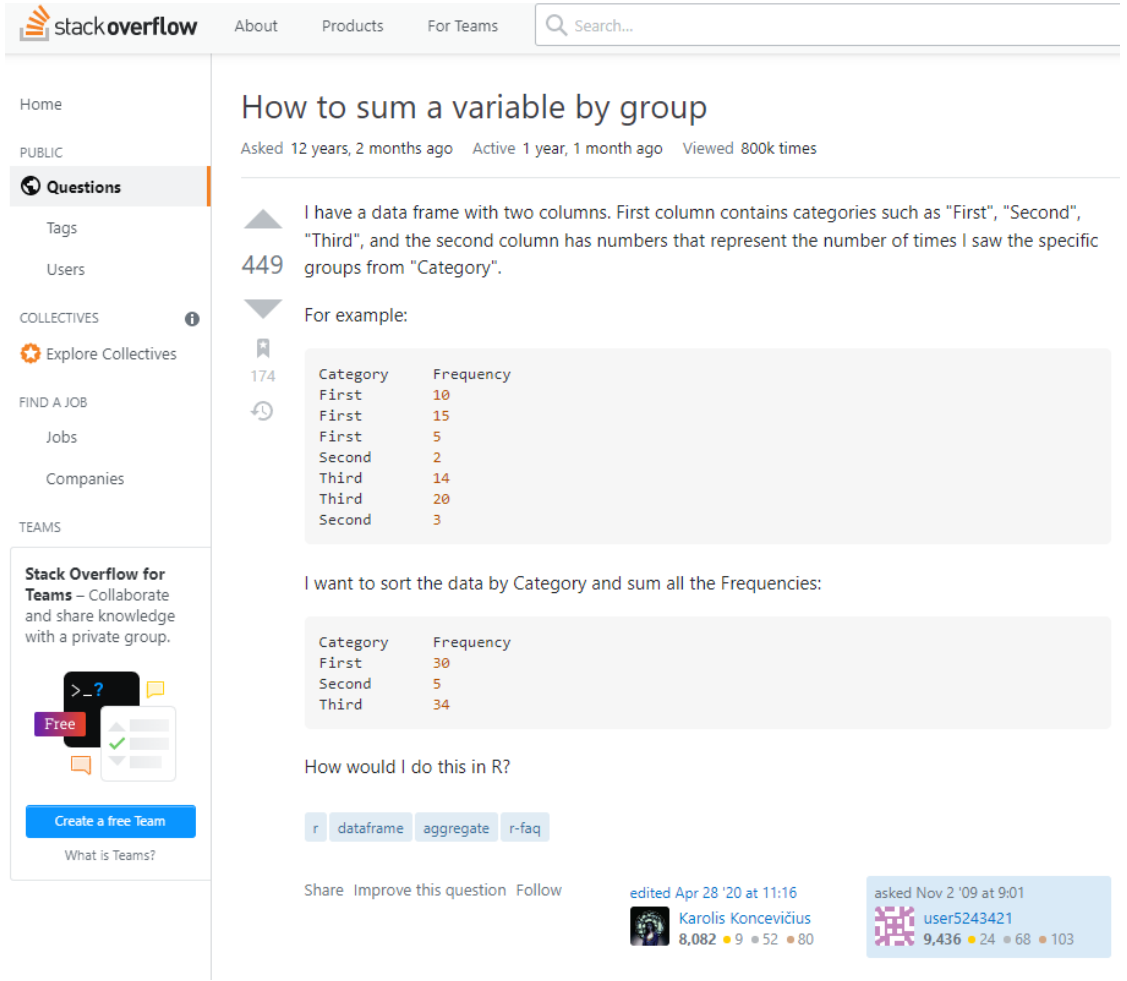


Your strongest ally



Your strongest ally





Your second strongest ally – the docs

```
> ?mean
```

mean {base} R Documentation

Arithmetic Mean

Description

Generic function for the (trimmed) arithmetic mean.

Usage

```
mean(x, ...)

## Default S3 method:
mean(x, trim = 0, na.rm = FALSE, ...)
```

Arguments

`x`	An `R` object. Currently there are methods for numeric/logical vectors and date, date-time and time interval objects. Complex vectors are allowed for `trim = 0`, only.
`trim`	the fraction (0 to 0.5) of observations to be trimmed from each end of `x` before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.
`na.rm`	a logical value indicating whether `NA` values should be stripped before the computation proceeds.
`...`	further arguments passed to or from other methods.

Value

If `trim` is zero (the default), the arithmetic mean of the values in `x` is computed, as a numeric or complex vector of length one. If `x` is not logical (coerced to numeric), numeric (including integer) or complex, `NA_real_` is returned, with a warning.

If `trim` is non-zero, a symmetrically trimmed mean is computed with a fraction of `trim` observations deleted from each end before the mean is computed.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

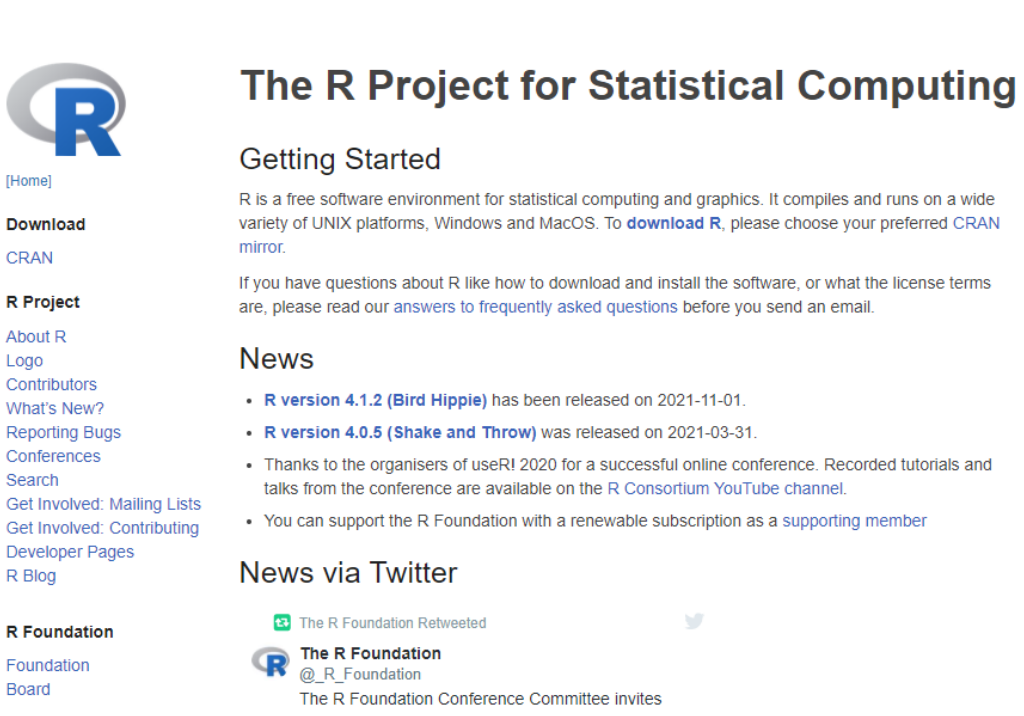
See Also

`weighted.mean`, `mean.POSIXct`, `colMeans` for row and column means.

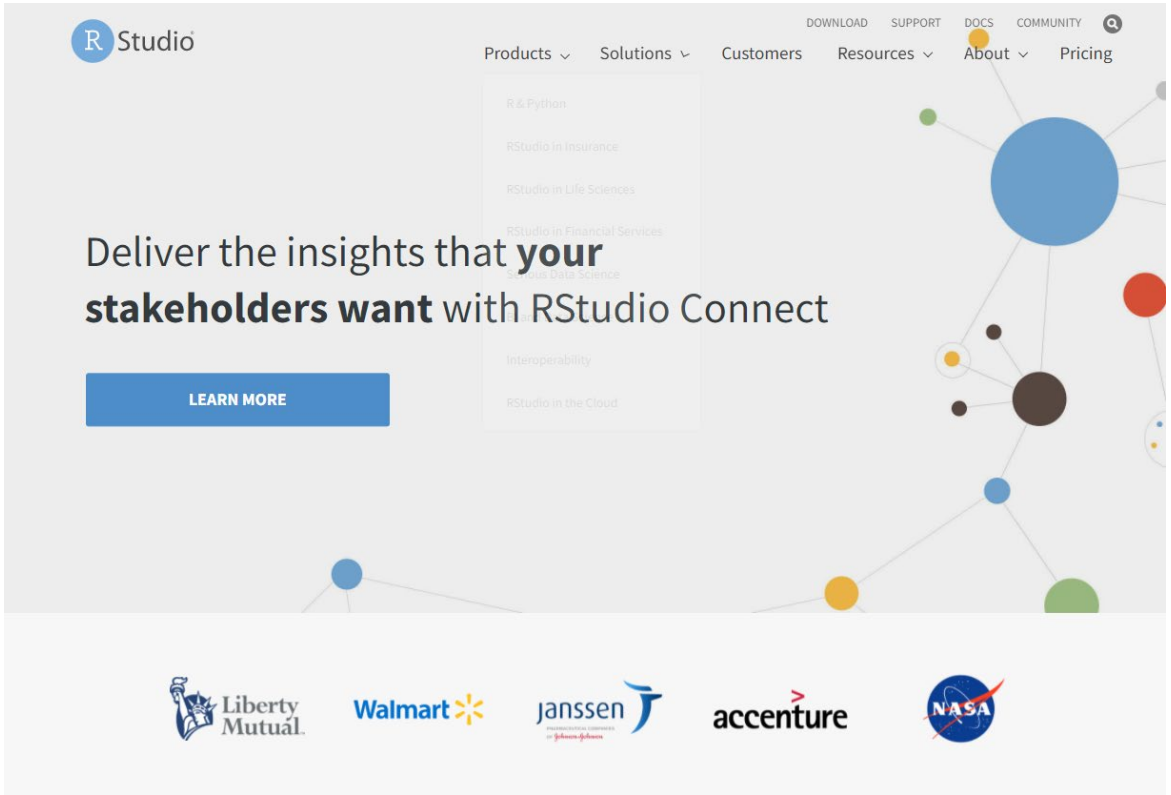
Examples

```
x <- c(0:10, 50)
xm <- mean(x)
c(xm, mean(x, trim = 0.10))
```

R and RStudio



https://www.r-project.org/



https://www.rstudio.com/