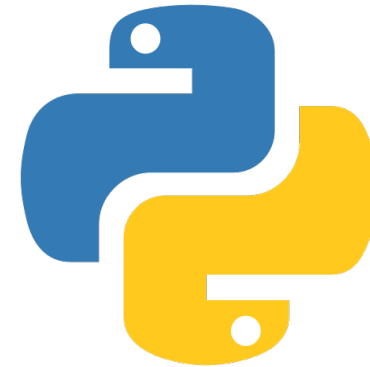


Software Choices:



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?

13																
1	EventID	EventType	StartDateTime	EndDateTime	EnteredOn	EventAgency	ParkingHeld	Borough	CommunityB	PolicePrecinct	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 HILL	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 STRE	Brooklyn	1, 2	108, 94	Television	Episodic seri	United State	11101, 11222		
8	66846	Shooting Per	#####	#####	#####	Mayor's Offi	8 AVENUE bo	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

```
1 average = function(numbers)
2 {
3   total = 0
4   for (i in 1:length(numbers))
5   {
6     total = total + numbers[i]
7   }
8   mean = total / length(numbers)
9
10  return(mean)
11 }
```

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

```
2
3 def average(numbers):
4
5     total = 0
6     for i in range(len(numbers)):
7         total = total + numbers[i]
8
9     mean = total / len(numbers)
10
11     return mean
12
```

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


Where to find data

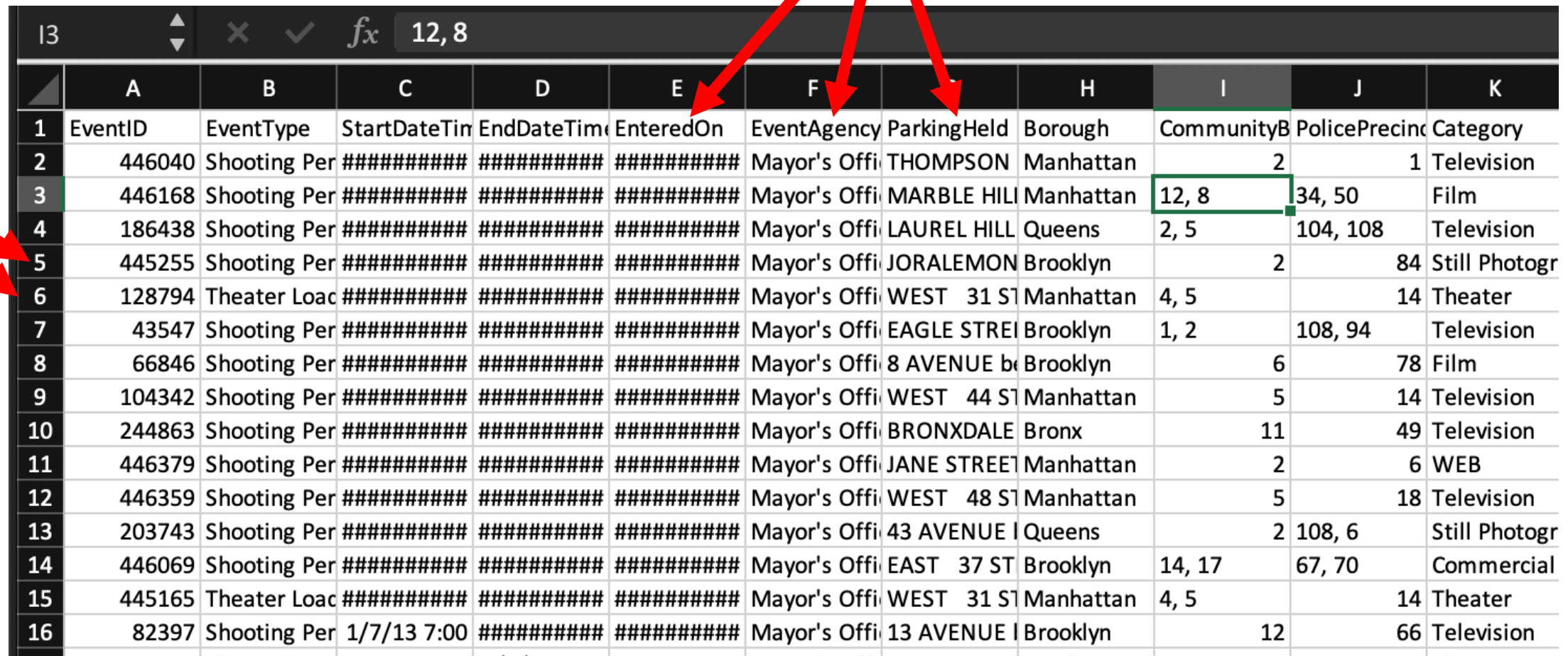


National Center for Health Statistics

Data organization

columns (variables)

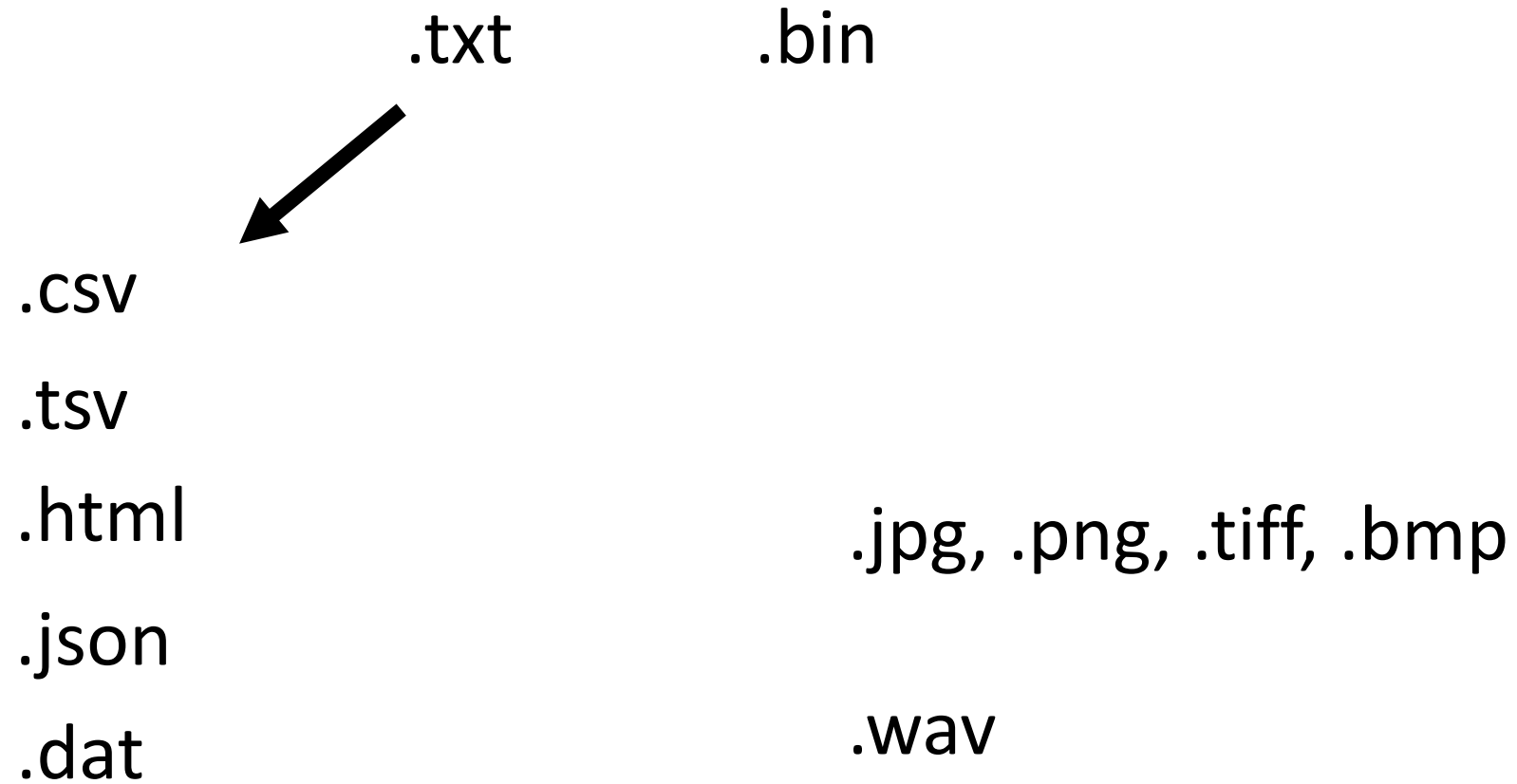
rows



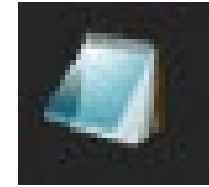
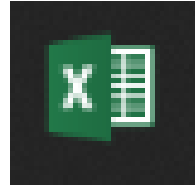
The image shows an Excel spreadsheet with a grid of data. Red arrows point from the text 'rows' to the first column (row numbers) and from the text 'columns (variables)' to the first row (column headers). The spreadsheet has a formula bar at the top showing '12,8' and a formula icon. The data is organized into columns labeled A through K, representing different variables. The rows are numbered 1 through 16. The data includes event IDs, event types, start and end dates, entered dates, event agencies, parking held, boroughs, community boards, police precincts, and categories.


	A	B	C	D	E	F	G	H	I	J	K
1	EventID	EventType	StartDateTin	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 STREET	Brooklyn	1, 2	108, 94	Television
8	66846	Shooting Per	#####	#####	#####	Mayor's Offi	8 AVENUE b	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



Common data formats



 song_duration_data.csv



	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



```
1 songname,isi,species,duration
2 au2219bf_song01,NA,BF,0.065193855
3 au2219bf_song01,0.118392041,BF,0.066985269
4 au2219bf_song01,0.079411785,BF,0.077438962
5 au2219bf_song01,0.078255302,BF,0.078890234
6 au2219bf_song01,0.061588352,BF,0.081634045
7 au2219bf_song01,0.009365239,BF,0.09465014
8 au2219bf_song01,0.037120814,BF,0.107190036
9 au2219bf_song01,0.012063697,BF,0.056554252
10 au2219bf_song01,0.015941315,BF,0.057892144
11 au2219bf_song01,0.014444691,BF,0.064196106
```


Common data formats



EXPLORE



```
<!DOCTYPE html>
<html lang="en" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://pu
r1.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://ogp.me/ns# rdfs: http://www.w3.org/2
000/01/rdf-schema# schema: http://schema.org/ sioc: http://rdfs.org/sioc/ns# sioc: http://rdfs.o
rg/sioc/types# skos: http://www.w3.org/2004/02/skos/core# xsd: http://www.w3.org/2001/XMLSchema# "
class="js cssanimations csstransitions objectfit object-fit wf-proximanova-n4-active wf-proximanova-
n3-active wf-proximanova-n5-active wf-trajanpro3-n4-active ng-scope wf-proximanova-n6-active wf-prox
imanova-n7-active wf-proximanova-n1-active wf-proximanova-i4-active wf-adobegaramondpro-i4-active wf
-adobegaramondpro-n4-active wf-active" data-ng-app="app">
  <head>...</head>
  <body class="cu-profile-admin node-home_landing_page cu-www path-frontpage page-node-type-home-1
anding-page has-glyphicons ng-scope banner-loaded mobile" data-ng-class="[global.controller, sear
h.widgetOpen ? 'search-open' : '', (global.controller == 'search' && gcse.status) ? 'search-ready'
: '', nav.mobileOpen ? 'mobile-nav-open' : '', page.overlayShown ? 'overlay-shown' : '', gsa.class
es]" data-breakpoint="{ 0:'mobile', 767:'break768', 991:'break992' }" data-ng-controller="global">
    <a id="top" href="#main-content" class="visually-hidden focusable skip-link"> Skip to main
content </a>
    <div class="dialog-off-canvas-main-canvas" data-off-canvas-main-canvas>...</div>
    <!-- Modal -->
    <div class="modal fade" id="modal" tabindex="-1" role="dialog" aria-labelledby="modal-heading"
aria-hidden="true">...</div>
    <script src="/content/sites/default/files/js/js_tXxb2tnFdZ95BxxExbt6b1WfwtcG05VG27uRje6_SUM.j
s"></script> == $0
    <script src="https://www.youtube.com/iframe api"></script>
    <script src="/content/sites/default/files/js/js_UuJgQXsE_CtH8uKkARMi-pfN1beyPgtAErrBypQ1DUE.j
s"></script>
    <script src="//s7.addthis.com/js/300/addthis_widget.js#pubid=ra-5321fb424ebb991b"></script>
    <div id="_atssh" style="visibility: hidden; height: 1px; width: 1px; position: absolute; top: -
9999px; z-index: 100000;">...</div>
    <style id="service-icons-0"></style>
    <script type="text/javascript"> var addthis_config = { ui_tabindex: 0 } </script>
  </body>
</html>
```


Your strongest ally



🔍 Search Google or type a URL



Your strongest ally

AboutProductsFor TeamsSearch...

Home

PUBLIC

Questions

Tags

Users

COLLECTIVES

Explore Collectives

FIND A JOB

Jobs

Companies

TEAMS

Stack Overflow for Teams – Collaborate and share knowledge with a private group.

Create a free Team

What is Teams?

How to rename a single column in a data.frame?

Asked 10 years, 4 months ago · Active 17 days ago · Viewed 1.1m times

406

I know if I have a data frame with more than 1 column, then I can use

```
colnames(x) <- c("col1", "col2")
```

to rename the columns. How to do this if it's just one column? Meaning a vector or data frame with only one column.

147

Example:

```
trSamp <- data.frame(sample(trainer$index, 10000))
head(trSamp )
# sample.trainer.index..10000.
# 1 5907862
# 2 2181266
# 3 7368504
# 4 1949790
# 5 3475174
# 6 6062879


ncol(trSamp)
# [1] 1
class(trSamp)
# [1] "data.frame"
class(trSamp[1])
# [1] "data.frame"
class(trSamp[,1])
# [1] "numeric"
colnames(trSamp)[2] <- "newname2"
# Error in names(x) <- value :
# 'names' attribute [2] must be the same length as the vector [1]
```

r

Share Improve this question Follow

edited Jan 5 at 21:31
Andrew Gillreath-Brown
4,388 · 3 · 6 · 33

asked Sep 23 '11 at 16:08
screechOwl
25.1k · 58 · 150 · 254

AboutProductsFor TeamsSearch...

Home

PUBLIC

Questions

Tags

Users

COLLECTIVES

Explore Collectives

FIND A JOB

Jobs

Companies

TEAMS

Stack Overflow for Teams – Collaborate and share knowledge with a private group.

Create a free Team

What is Teams?

How to sum a variable by group

Asked 12 years, 2 months ago · Active 1 year, 1 month ago · Viewed 800k times

449

I have a data frame with two columns. First column contains categories such as "First", "Second", "Third", and the second column has numbers that represent the number of times I saw the specific groups from "Category".

For example:

Category	Frequency
First	10
First	15
First	5
Second	2
Third	14
Third	20
Second	3

I want to sort the data by Category and sum all the Frequencies:

Category	Frequency
First	30
Second	5
Third	34

How would I do this in R?

r dataframe aggregate r-faq

Share Improve this question Follow

edited Apr 28 '20 at 11:16
Karolis Koncevičius
8,082 · 9 · 52 · 80

asked Nov 2 '09 at 9:01
user5243421
9,436 · 24 · 68 · 103

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



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing Lists](#)

[Get Involved: Contributing](#)

[Developer Pages](#)

[R Blog](#)

R Foundation

[Foundation](#)

[Board](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.1.2 (Bird Hippie)** has been released on 2021-11-01.
- **R version 4.0.5 (Shake and Throw)** was released on 2021-03-31.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the [R Consortium YouTube channel](#).
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

News via Twitter

The R Foundation Retweeted



The R Foundation

@_R_Foundation

The R Foundation Conference Committee invites

A screenshot of the RStudio Connect website. The header includes the RStudio logo and navigation links: Products, Solutions, Customers, Resources, About, and Pricing. A sidebar on the left lists various solutions like R & Python, RStudio in Insurance, etc. The main content area features a large headline: "Deliver the insights that **your stakeholders want** with RStudio Connect", followed by a blue "LEARN MORE" button. The footer displays logos of partner organizations: Liberty Mutual, Walmart, Janssen, Accenture, and NASA. The background has a network diagram with blue and red nodes.

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

<https://www.rstudio.com/>