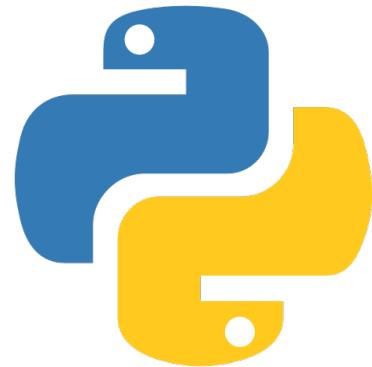


Software Choices:



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?

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|----|---------|--------------|---------------|-------------|-----------|--------------|--------------|-----------|------------|----------------|---------------|---------------|--------------|--------------|---|
| 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 | Bronoklyn | 2 | 84 | Still Photogr | 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 STREET | Brooklyn | 1, 2 | 108, 94 | Television | Episodic seri | United State | 11101, 11222 | |
| 8 | 66846 | Shooting Per | ##### | ##### | ##### | Mayor's Offi | 8 AVENUE | 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 | Queens | 2 | 108, 6 | Still Photogr | 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 | 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
13
```

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

Where to find data



National Center for Health Statistics

Data organization

columns (variables)

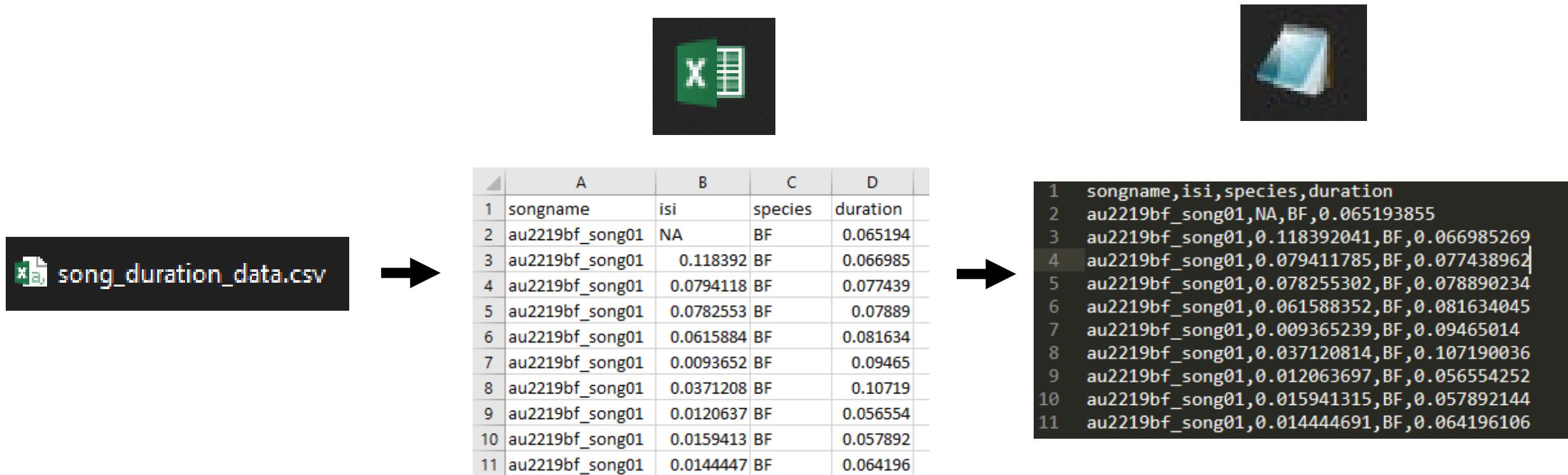
rows

| I3 | A | B | C | D | E | F | G | H | I | J | K |
|----|---------|-----------|--------------|-------------|-----------|-------------|-------------|--------------|--------------|----------------|----------|
| 1 | EventID | EventType | StartDateTim | EndDateTime | EnteredOn | EventAgency | ParkingHeld | Borough | CommunityB | PolicePrecinct | Category |
| 2 | 446040 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | THOMPSON | Manhattan | 2 |
| 3 | 446168 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | MARBLE HILL | Manhattan | 34, 50 |
| 4 | 186438 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | LAUREL HILL | Queens | 104, 108 |
| 5 | 445255 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | JORALEMON | Brooklyn | 84 |
| 6 | 128794 | Theater | Load | ##### | ##### | ##### | ##### | Mayor's Offi | WEST 31 ST | Manhattan | 14 |
| 7 | 43547 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | EAGLE STREET | Brooklyn | 108, 94 |
| 8 | 66846 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | 8 AVENUE | Brooklyn | 78 |
| 9 | 104342 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | WEST 44 ST | Manhattan | 14 |
| 10 | 244863 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | BRONXDALE | Bronx | 49 |
| 11 | 446379 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | JANE STREET | Manhattan | 6 |
| 12 | 446359 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | WEST 48 ST | Manhattan | 18 |
| 13 | 203743 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | 43 AVENUE | Queens | 108, 6 |
| 14 | 446069 | Shooting | Per | ##### | ##### | ##### | ##### | Mayor's Offi | EAST 37 ST | Brooklyn | 67, 70 |
| 15 | 445165 | Theater | Load | ##### | ##### | ##### | ##### | Mayor's Offi | WEST 31 ST | Manhattan | 14 |
| 16 | 82397 | Shooting | Per | 1/7/13 7:00 | ##### | ##### | ##### | Mayor's Offi | 13 AVENUE | Brooklyn | 66 |

Common data formats

| | | |
|-------|------|-------------------------|
| | .txt | .bin |
| .CSV | | |
| .tsv | | |
| .html | | .jpg, .png, .tiff, .bmp |
| .json | | |
| .dat | | .wav |

Common data formats



Common data formats

The screenshot shows the homepage of the Columbia University Mailman School of Public Health. At the top, there's a blue header bar with the Columbia University logo and a search icon. Below the header is a large collage of images featuring students, faculty, and buildings. Overlaid on the collage is the text "Mailman School of Public Health". A blue banner across the middle of the page reads "Building a Just and Healthy World for 100 Years". Below this banner are two smaller images: one showing people working at tables and another showing a woman speaking into a microphone. At the bottom of the page, there's a section titled "EXPLORE" with two images: one of a modern building at night and another of a classical building at night.

The screenshot shows the "Elements" tab of a browser developer tools window. It displays the raw HTML code of the page. The code includes various CSS classes and IDs, some of which are highlighted in blue and orange, likely indicating they are selected or being inspected. The code is heavily commented with "<!--" and "-->" blocks, and it includes several script tags pointing to external files like "/content/sites/default/files/js/js_txXb2tnFdz95BxxExbt6b1WFwtcG05VG27uRje6_SUM.js" and "https://www.youtube.com/embed/api".</p>

```
<!DOCTYPE html>
<html lang="en" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://ogp.me/ns# rdfs: http://www.w3.org/2000/01/rdf-schema# schema: http://schema.org/ sioc: http://rdfs.org/sioc/ns# sioc: http://rdfs.org/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-proximanova-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 landing-page has-glyphicons ng-scope banner-loaded mobile" data-ng-class="[global.controller, search.widgetOpen ? 'search-open' : '', (global.controller == 'search' && gcse.status) ? 'search-ready' : '', nav.mobileOpen ? 'mobile-nav-open' : '', page.overlayShown ? 'overlay-shown' : '', gsa.classes]" 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.js">...</script> == $0
    <script src="https://www.youtube.com/embed/api"></script>
    <script src="/content/sites/default/files/js/js_UuJgOxsE_CtH8uKkARMi-pfN1beyPqtAErrBypQ1DUE.js">...</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

stack overflow

About Products For Teams Search...

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.

>?

Free

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

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

406 `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 asked Sep 23 '11 at 16:08 screechOwl

4,388 3 6 33 25.1k 58 150 254

stack overflow

About Products For Teams Search...

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.

>?

Free

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

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".

449 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 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

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

R Studio

Products ▾ Solutions ▾ Customers Resources ▾ DOCS COMMUNITY About ▾ Pricing

Deliver the insights that **your stakeholders want** with RStudio Connect

LEARN MORE

R & Python
RStudio in Insurance
RStudio in Life Sciences
RStudio in Financial Services
RStudio Data Science
Interoperability
RStudio in the Cloud

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