

# Package: covdata (via r-universe)

October 23, 2024

**Type** Package

**Title** COVID-19 Data

**Version** 1.01

**Description** COVID-19 related data from the ECDC, the COVID-19 Tracking Project, the New York Times, the Human Mortality Database, and Apple. Packaged for R.

**Depends** R ( $\geq 2.10$ )

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.3

**URL** <https://github.com/kjhealy/covdata>

**BugReports** <https://github.com/kjhealy/covdata/issues>

**Imports** dplyr, magrittr, tidyr

**Suggests** drat, ggforce, ggrepel, lubridate, knitr, rmarkdown, scales, skimr, slider, tidyverse, testthat

**VignetteBuilder** knitr

**Repository** <https://kjhealy.r-universe.dev>

**RemoteUrl** <https://github.com/kjhealy/covdata>

**RemoteRef** HEAD

**RemoteSha** 0626c6522c38f1ac0c74453267dd5779a88ca2c6

## Contents

apple_mobility . . . . .	2
cdc_catchments . . . . .	4
cdc_deaths_by_age . . . . .	6

cdc_deaths_by_sex . . . . .	8
cdc_deaths_by_state . . . . .	10
cdc_deaths_by_week . . . . .	12
countries . . . . .	14
covnat_daily . . . . .	15
covnat_weekly . . . . .	16
covus . . . . .	18
covus_ethnicity . . . . .	21
covus_race . . . . .	22
fmt_nc . . . . .	24
fmt_nr . . . . .	25
MMWRweek2Date . . . . .	26
MMWRweekday . . . . .	27
mmwr_week_to_date . . . . .	28
nchs_sas . . . . .	29
nchs_wdc . . . . .	31
nchs_wss . . . . .	32
nssp_covid_er_nat . . . . .	35
nssp_covid_er_reg . . . . .	37
nytcovcounty . . . . .	38
nytcovstate . . . . .	40
nytcovus . . . . .	42
nytexcess . . . . .	43
start_date . . . . .	45
stmf . . . . .	46
stmf_country_years . . . . .	49
tabular . . . . .	50
uspop . . . . .	51
%nin% . . . . .	53

<b>Index</b>	<b>54</b>
--------------	-----------

---

apple_mobility	<i>Apple Mobility Data</i>
----------------	----------------------------

---

## Description

Data from Apple Maps on relative changes in mobility in various cities and countries.

## Usage

apple\_mobility

**Format**

A data frame with 2,254,515 rows and 7 variables:

`country` character Country name (not provided for all countries)

`sub_region` character Subregion names

`subregion_and_city` character Subregion and city names

`geo_type` character Type geographical unit. Values: city, country/region, sub-region

`transportation_type` character Mode of transport. Values: driving, transit, or walking

`date` double Date in yyyy-mm-dd format

`score` double Activity score. Indexed to 100 on the first date of observation for a given mode of transport.

**Details**

Table: Data summary

Name	apple_mobility
Number of rows	2254515
Number of columns	7
<hr/>	
Column type frequency:	
Date	1
character	5
numeric	1
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-01-13	2022-04-12	2021-02-26	819

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
country	0	1	5	20	0	63	0
sub_region	0	1	4	46	0	606	0
subregion_and_city	0	1	4	46	0	853	0
geo_type	0	1	4	14	0	3	0
transportation_type	0	1	7	7	0	3	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
score	608041	0.73	122.59	66.81	2.43	83.79	113.72	148.8	2228.83	

Data made available by Apple, Inc. at <https://www.apple.com/covid19/mobility>, showing relative volume of directions requests per country/region or city compared to a baseline volume on January 13th, 2020. Apple defines the day as midnight-to-midnight, Pacific time. Cities represent usage in greater metropolitan areas and are stably defined during this period. In many countries/regions and cities, relative volume has increased since January 13th, consistent with normal, seasonal usage of Apple Maps. Day of week effects are important to normalize as you use this data. Data that is sent from users' devices to the Apple Maps service is associated with random, rotating identifiers so Apple does not have a profile of individual movements and searches. Apple Maps has no demographic information about its users, and so cannot make any statements about the representativeness of its usage against the overall population.

### Author(s)

Kieran Healy

### Source

<https://www.apple.com/covid19/mobility>

### References

See <https://www.apple.com/covid19/mobility> for detailed terms of use.

---

cdc_catchments	<i>CDC surveillance network and network catchment area</i>
----------------	--

---

### Description

What the CDC surveillance network covers

### Usage

cdc\_catchments

### Format

A data frame with 17 rows and 2 variables:

**name** character Network name

**area** character Area

**Details**

Table: Data summary

Name	cdc_catchments
Number of rows	17
Number of columns	2
<hr/>	
Column type frequency:	
character	2
<hr/>	
Group variables	None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
name	0	1	3	9	0	3	0
area	0	1	4	14	0	15	0

The Coronavirus Disease 2019 (COVID-19)-Associated Hospitalization Surveillance Network (COVID-NET) conducts population-based surveillance for laboratory-confirmed COVID-19-associated hospitalizations in children (persons younger than 18 years) and adults. The current network covers nearly 100 counties in the 10 Emerging Infections Program (EIP) states (CA, CO, CT, GA, MD, MN, NM, NY, OR, and TN) and four additional states through the Influenza Hospitalization Surveillance Project (IA, MI, OH, and UT). The network represents approximately 10% of US population (~32 million people). Cases are identified by reviewing hospital, laboratory, and admission databases and infection control logs for patients hospitalized with a documented positive SARS-CoV-2 test. Data gathered are used to estimate age-specific hospitalization rates on a weekly basis and describe characteristics of persons hospitalized with COVID-19. Laboratory confirmation is dependent on clinician-ordered SARS-CoV-2 testing. Therefore, the unadjusted rates provided are likely to be underestimated as COVID-19-associated hospitalizations can be missed due to test availability and provider or facility testing practices. COVID-NET hospitalization data are preliminary and subject to change as more data become available. All incidence rates are unadjusted. Please use the following citation when referencing these data: “COVID-NET: COVID-19-Associated Hospitalization Surveillance Network, Centers for Disease Control and Prevention. WEBSITE. Accessed on DATE”.

name	area
COVID-NET	Entire Network
EIP	California
EIP	Colorado
EIP	Connecticut
EIP	Entire Network
EIP	Georgia

EIP	Maryland
EIP	Minnesota
EIP	New Mexico
EIP	New York
EIP	Oregon
EIP	Tennessee
IHSP	Entire Network
IHSP	Iowa
IHSP	Michigan
IHSP	Ohio
IHSP	Utah

**Author(s)**

Kieran Healy

**Source**

Courtesy of Bob Rudis's cdcovidview package

**References**

<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>

---

cdc\_deaths\_by\_age      *CDC Surveillance Network Death Counts by Age*

---

**Description**

Provisional Death Counts for Coronavirus Disease (COVID-19)

**Usage**

cdc\_deaths\_by\_age

**Format**

A data frame with 12 rows and 10 variables:

`data_as_of` date When the data were most recently recorded

`age_group` character Age range

`start_week` date Start week

`end_week` date End week

`covid_deaths` integer COLUMN\_DESCRIPTION

`total_deaths` integer COLUMN\_DESCRIPTION

```

percent_expected_deaths double COLUMN_DESCRIPTION
pneumonia_deaths integer COLUMN_DESCRIPTION
pneumonia_and_covid_deaths integer COLUMN_DESCRIPTION
all_influenza_deaths_j09_j11 integer COLUMN_DESCRIPTION

```

## Details

Table: Data summary

Name	cdc_deaths_by_age
Number of rows	12
Number of columns	10
<hr/>	
Column type frequency:	
Date	3
character	1
numeric	6
<hr/>	
Group variables	None

### Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2020-04-30	2020-04-30	2020-04-30	1
start_week	0	1	2020-02-01	2020-02-01	2020-02-01	1
end_week	0	1	2020-04-25	2020-04-25	2020-04-25	1

### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
age_group	0	1	5	10	0	12	0

### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p
covid_deaths	0	1	5753.50	9877.31	2.00	30.25	1211.
total_deaths	0	1	118897.67	202377.07	712.00	5675.25	28460.
percent_expected_deaths	0	1	0.97	0.00	0.97	0.97	0.
pneumonia_deaths	0	1	10454.17	18036.25	33.00	109.00	1799.
pneumonia_and_covid_deaths	0	1	2550.17	4387.93	0.00	12.50	491.
all_influenza_deaths_j09_j11	0	1	970.17	1618.90	11.00	40.75	358.

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdcovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)). Please see the indicated reference for all the caveats and precise meanings for each field.

### Author(s)

Kieran Healy

### Source

Courtesy of Bob Rudis's `cdcovidview` package

### References

<https://data.cdc.gov/api/views/hc4f-j6nb/rows.csv?accessType=DOWNLOAD&bom=true&format=true>

---

cdc\_deaths\_by\_sex      *CDC provisional death counts by sex*

---

### Description

Provisional Death Counts for Coronavirus Disease (COVID-19)

### Usage

`cdc_deaths_by_sex`

### Format

A data frame with 3 rows and 10 variables:

`data_as_of` date Date most recently updated

`sex` character Sex

`start_week` date Beginning week

`end_week` date Ending week

`covid_deaths` integer COVID deaths

`total_deaths` integer Total deaths

`percent_expected_deaths` double COLUMN\_DESCRIPTION

`pneumonia_deaths` integer COLUMN\_DESCRIPTION

`pneumonia_and_covid_deaths` integer COLUMN\_DESCRIPTION

`all_influenza_deaths_j09_j11` integer COLUMN\_DESCRIPTION



**Details**

Table: Data summary

Name	cdc_deaths_by_sex
Number of rows	3
Number of columns	10
<hr/>	
Column type frequency:	
Date	3
character	1
numeric	6
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2020-04-30	2020-04-30	2020-04-30	1
start_week	0	1	2020-02-01	2020-02-01	2020-02-01	1
end_week	0	1	2020-04-25	2020-04-25	2020-04-25	1

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
sex	0	1	4	7	0	3	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
covid_deaths	0	1	11507.33	10231.40	1.00	7470.50	14940.00	22450.00	14940.00
total_deaths	0	1	237795.00	206241.06	25.00	172555.00	345080.00	517635.00	517635.00
percent_expected_deaths	0	1	0.97	0.00	0.97	0.97	0.97	0.97	0.97
pneumonia_deaths	0	1	20908.33	18248.40	1.00	14545.00	29090.00	43635.00	43635.00
pneumonia_and_covid_deaths	0	1	5100.33	4559.67	1.00	3258.00	6516.00	9774.00	9774.00
all_influenza_deaths_j09_j11	0	1	1940.33	1682.21	0.00	1416.00	2832.00	4248.00	4248.00

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdccovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)). Please see the indicated reference for all the caveats and precise meanings for each field.

**Author(s)**

Kieran Healy

**Source**

Courtesy of Bob Rudis's cdccovidview package

**References**

<https://data.cdc.gov/api/views/hc4f-j6nb/rows.csv?accessType=DOWNLOAD&bom=true&format=true>

---

cdc\_deaths\_by\_state    *CDC provisional death counts by state*

---

**Description**

CDC Surveillance Network provisional death counts

**Usage**

cdc\_deaths\_by\_state

**Format**

A data frame with 53 rows and 10 variables:

data\_as\_of date Date most recently updated

state character State name

start\_week date Start week

end\_week double End week

covid\_deaths integer COVID Deaths

total\_deaths integer Total deaths

percent\_expected\_deaths double COLUMN\_DESCRIPTION

pneumonia\_deaths integer COLUMN\_DESCRIPTION

pneumonia\_and\_covid\_deaths integer COLUMN\_DESCRIPTION

all\_influenza\_deaths\_j09\_j11 integer COLUMN\_DESCRIPTION

**Details**

Table: Data summary

Name	cdc_deaths_by_state
Number of rows	53
Number of columns	10
<hr/>	
Column type frequency:	
Date	3
character	1
numeric	6
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2020-04-30	2020-04-30	2020-04-30	1
start_week	0	1	2020-02-01	2020-02-01	2020-02-01	1
end_week	0	1	2020-04-25	2020-04-25	2020-04-25	1

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1	4	20	0	53	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50
covid_deaths	6	0.89	735.02	1801.11	0	54.50	153.00
total_deaths	0	1.00	13557.43	13996.83	856	3813.00	10721.00
percent_expected_deaths	0	1.00	0.93	0.27	0	0.86	0.95
pneumonia_deaths	0	1.00	1197.26	1453.17	41	277.00	769.00
pneumonia_and_covid_deaths	10	0.81	355.81	759.51	0	30.50	65.00
all_influenza_deaths_j09_j11	3	0.94	116.58	142.24	14	30.50	87.50

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdccovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET. Please see the indicated reference for all the caveats and precise meanings for each field. ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)).

**Author(s)**

Kieran Healy

**References**

<https://data.cdc.gov/api/views/hc4f-j6nb/rows.csv?accessType=DOWNLOAD&bom=true&format=true>

---

cdc\_deaths\_by\_week     *CDC Provisional death counts by week*

---

**Description**

Provisional Death Counts for Coronavirus Disease (COVID-19)

**Usage**

cdc\_deaths\_by\_week

**Format**

A data frame with 13 rows and 10 variables:

data\_as\_of date When the data were most recently recorded

start\_week date Start week

end\_week double End week

covid\_deaths integer COVID deaths

total\_deaths integer Total deaths

percent\_expected\_deaths double COLUMN\_DESCRIPTION

pneumonia\_deaths integer COLUMN\_DESCRIPTION

pneumonia\_and\_covid\_deaths integer COLUMN\_DESCRIPTION

all\_influenza\_deaths\_j09\_j11 integer COLUMN\_DESCRIPTION

pneumonia\_influenza\_and\_covid\_19\_deaths integer COLUMN\_DESCRIPTION

**Details**

Table: Data summary

Name	cdc_deaths_by_week
Number of rows	13
Number of columns	10

---

Column type frequency:

Date	3
numeric	7
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2020-04-30	2020-04-30	2020-04-30	1
start_week	0	1	2020-02-01	2020-04-25	2020-03-14	13
end_week	0	1	2020-02-01	2020-04-25	2020-03-14	13

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0
covid_deaths	0	1	2655.46	4194.37	0.00
total_deaths	0	1	54875.85	9864.46	24387.00
percent_expected_deaths	0	1	0.97	0.17	0.45
pneumonia_deaths	0	1	4825.00	2217.19	2219.00
pneumonia_and_covid_deaths	0	1	1177.00	1863.76	0.00
all_influenza_deaths_j09_j11	0	1	447.77	156.19	58.00
pneumonia_influenza_and_covid_19_deaths	0	1	6690.23	4292.62	3553.00

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdccovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)). Please see the indicated reference for all the caveats and precise meanings for each field.

**Author(s)**

Kieran Healy

**Source**

Courtesy of Bob Rudis's `cdccovidview` package

**References**

<https://data.cdc.gov/api/views/hc4f-j6nb/rows.csv?accessType=DOWNLOAD&bom=true&format=true>

---

`countries` *Country Names and ISO codes*

---

### Description

Convenience table of country names and their abbreviated names

### Usage

```
countries
```

### Format

A data frame with 213 rows and 4 variables:

```
cname character Country name
iso3 character ISO 3 designation
iso2 character ISO 2 designation
continent Continent
```

### Details

Table: Data summary

Name	dplyr::ungroup(countries)
Number of rows	213
Number of columns	4
<hr/>	
Column type frequency:	
character	4
<hr/>	
Group variables	None

### Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
cname	0	1.00	4	42	0	213	0
iso3	0	1.00	3	3	0	213	0
iso2	2	0.99	2	2	0	211	0
continent	0	1.00	4	13	0	6	0

Produced from the ECDC tables in the covdata package.

**Author(s)**

Kieran Healy

**References**

ISO 2: [https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2) ISO 3: [https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-3](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3)

covnat\_daily

*Daily international COVID-19 cases and deaths for 2020***Description**

A dataset containing daily national-level ECDC data on COVID-19. Archived as of December 14th 2020. ECDC switched to a weekly reporting schedule for the COVID-19 situation worldwide and in the EU/EEA and the UK on 17 December 2020. Daily updates have been discontinued from 14 December 2020.

**Usage**

covnat\_daily

**Format**

A tibble with 61,836 rows and 8 columns

**date** date in YYYY-MM-DD format

**cname** Name of country (character)

**iso3** ISO3 country code (character)

**cases** N reported COVID-19 cases for this day

**deaths** N reported COVID-19 deaths for this day

**pop** Country population from Eurostat or UN data

**cu\_cases** Cumulative N reported COVID-19 cases up to and including this day

**cu\_deaths** Cumulative N reported COVID-19 deaths up to and including this day

**Details**

Table: Data summary

Name	dplyr::ungroup(covnat_dai...
Number of rows	61836
Number of columns	8

---

Column type frequency:

Date	1
character	2
numeric	5
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2019-12-31	2020-12-14	2020-07-21	350

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
cname	0	1	4	42	0	213	0
iso3	0	1	3	3	0	213	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
cases	0	1	1156.33	6782.63	-8261	0	15	27	27
deaths	0	1	26.08	131.29	-1918	0	0	4	4
pop	59	1	40987698.23	153129379.34	815	1293120	7169456	2851582	2851582
cu_cases	0	1	100686.99	607743.06	0	129	2055	2465	2465
cu_deaths	0	1	3104.89	15545.84	0	1	42	46	46

**Source**

<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-c>

---

covnat_weekly	<i>Weekly International COVID-19 cases and deaths, current as of Sunday, January 22, 2023</i>
---------------	---

---

**Description**

A dataset containing weekly national-level ECDC data on COVID-19

**Usage**

covnat\_weekly



**Format**

A tibble with 4,966 rows and 11 columns

**date** date in YYYY-MM-DD format

**year\_week** Year and week of reporting (character, YYYY-WW)

**cname** Name of country (character)

**pop** Country population from Eurostat or UN data

**iso3** ISO3 country code (character)

**cases** N reported COVID-19 cases for this week

**deaths** N reported COVID-19 deaths for this week

**cu\_cases** Cumulative N reported COVID-19 cases up to and including this week

**cu\_deaths** Cumulative N reported COVID-19 deaths up to and including this week

**r14\_cases** 14-day notification rate of reported COVID-19 cases per 100,000 population

**r14\_deaths** 14-day notification rate of reported COVID-19 cases per 100,000 population

**Details**

Table: Data summary

Name	dplyr::ungroup(covnat_wee...
Number of rows	4966
Number of columns	11
<hr/>	
Column type frequency:	
Date	1
character	3
numeric	7
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2019-12-30	2023-01-09	2021-07-05	159

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
year_week	0	1.00	7	7	0	159	0
cname	0	1.00	5	14	0	31	0
iso3	196	0.96	3	3	0	30	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50
pop	0	1.00	31613614.13	85253844.55	39055	2108977.00	6916548.00
cases	222	0.96	77511.62	374657.80	0	1127.00	5487.00
deaths	279	0.94	514.14	2005.64	0	8.00	46.00
cu_cases	222	0.96	4188407.63	16969793.99	0	43400.25	485047.50
cu_deaths	279	0.94	44362.78	142967.65	0	651.00	6268.00
r14_cases	263	0.95	557.34	1044.46	0	51.61	216.74
r14_deaths	321	0.94	34.08	50.74	0	3.81	14.21

**Source**

<http://ecdc.europa.eu/>

---

covus	<i>COVID-19 data for the USA, current as of Sunday, January 22, 2023</i>
-------	--

---

**Description**

A dataset containing US state-level data on COVID-19

**Usage**

covus

**Format**

A tibble with 664,960 rows and 7 columns

**date** Date in YYYY-MM-DD format (date)

**state** Two letter State abbreviation (character)

**fips** State FIPS code (character)

**data\_quality\_grade** character Data quality as assessed by COVID Tracking Project staff

**measure** Outcome measure for this date

**count** Count of measure

**measure\_label** character Outcome measure, suitable for use as a plot label

**Details**

Table: Data summary

Name	covus
Number of rows	664960
Number of columns	7
<hr/>	
Column type frequency:	
Date	1
character	4
logical	1
numeric	1
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-01-13	2021-03-07	2020-09-03	420

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1	2	2	0	56	0
fips	0	1	2	2	0	56	0
measure	0	1	5	30	0	31	0
measure_label	0	1	6	54	0	32	0

**Variable type: logical**

skim_variable	n_missing	complete_rate	mean	count
data_quality_grade	664960	0	NaN	:

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
count	434365	0.35	387436.8	1638507	0	498	7782	134223	49646014	

The measures tracked by the COVID tracking project are as follows:

measure	measure_label
positive	Positive Tests
probable_cases	Probable Cases
negative	Negative Tests
pending	Pending Tests
hospitalized_currently	Currently Hospitalized
hospitalized_cumulative	Cumulative Hospitalized
in_icu_currently	Currently in ICU
in_icu_cumulative	Cumulative in ICU
on_ventilator_currently	Currently on Ventilator
on_ventilator_cumulative	Cumulative on Ventilator
recovered	Recovered
death	Deaths
hospitalized_discharged	Total Discharged from Hospital
total_tests_viral	Total number of PCR tests performed
positive_tests_viral	Total number of positive PCR tests
negative_tests_viral	Total number of negative PCR tests
positive_cases_viral	Total number of positive cases measured with PCR tests
death_confirmed	Deaths Confirmed
death_probable	Deaths Probable
total_test_encounters_viral	Total Test Encounters (PCR)
total_tests_people_viral	Total PCR Tests (People)
total_tests_antibody	Total Antibody Tests
positive_tests_antibody	Positive Antibody Tests
negative_tests_antibody	Total number of negative antibody tests
negative_tests_antibody	Negative Antibody Tests
total_tests_people_antibody	Total Antibody Tests (People)
positive_tests_people_antibody	Positive Antibody Tests (People)
negative_tests_people_antibody	Negative Antibody Tests (People)
total_tests_people_antigen	Total Antigen Tests (People)
positive_tests_people_antigen	Positive Antigen Tests (People)
total_tests_antigen	Total Antigen Tests
positive_tests_antigen	Positive Antigen Tests

Not all measures are reported by all states. The **positive**, **negative**, **death**, **death\_confirmed**, **probable\_cases** and **death\_probable** measures are *cumulative* counts. **death\_confirmed** is the total number deaths of individuals with COVID-19 infection confirmed by a laboratory test. In states where the information is available, it tracks only those laboratory-confirmed deaths where COVID also contributed to the death according to the death certificate. **death\_probable** is the total number of deaths where COVID was listed as a cause of death and there is not a laboratory test confirming COVID-19 infection.

For further information on the COVID Tracking Project's measures, see <https://covidtracking.com/about-data/data-definitions>

### Source

The COVID-19 Tracking Project <https://covidtracking.com>

---

<code>covus_ethnicity</code>	<i>COVID-19 case and death counts for the USA by Hispanic/Non-Hispanic ethnicity and state current as of Sunday, January 22, 2023</i>
------------------------------	---

---

### Description

The COVID Racial Data Tracker advocates for, collects, publishes, and analyzes racial data on the pandemic across the United States. It's a collaboration between the COVID Tracking Project and the Boston University Center for Antiracist Research.

### Usage

```
covus_ethnicity
```

### Format

A tibble with 15,960 rows and 7 columns

`date` date Data reported as of this date

`state` character State

`group` character Ethnic group

`cases` integer Total cases, count

`deaths` integer Total deaths, count

`hosp` integer Total hospitalizations, count

### Details

Table: Data summary

Name	<code>covus_ethnicity</code>
Number of rows	15960
Number of columns	7
<hr/>	
Column type frequency:	
Date	1
character	2
numeric	4
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-04-12	2021-03-07	2020-09-23	95

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1	2	2	0	56	0
group	0	1	7	12	0	3	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
cases	3080	0.81	73357.18	166184.31	0	5529	21920.5	70265.5	261944.0
deaths	3144	0.80	1645.64	3463.93	-1	63	291.5	1401.0	3200.0
hosp	11662	0.27	5079.37	8831.52	0	556	1556.0	4959.5	5600.0
tests	14271	0.11	892566.44	2376098.22	0	58933	224156.0	537668.0	2163300.0

The `group` variable is coded as "Hispanic", "Non-Hispanic", or "Unknown". Hispanics may be of any race. State-level counts should be handled with care, given the widely varying population distribution of people of different ethnic backgrounds by state.

**Author(s)**

Kieran Healy

**Source**

<https://covidtracking.com/race>

---

<code>covus_race</code>	<i>COVID-19 case and death counts for the USA by race and state current as of Sunday, January 22, 2023</i>
-------------------------	--

---

**Description**

The COVID Racial Data Tracker advocates for, collects, publishes, and analyzes racial data on the pandemic across the United States. It's a collaboration between the COVID Tracking Project and the Boston University Center for Antiracist Research.

**Usage**

`covus_race`

**Format**

A tibble with 47,880 rows and 7 columns

**date** date Data reported as of this date

**state** character State

**group** character Racial group

**cases** integer Total cases, count

**deaths** integer Total deaths, count

**hosp** integer Total hospitalizations, count

**Details**

Table: Data summary

Name	covus_race
Number of rows	47880
Number of columns	7
<hr/>	
Column type frequency:	
Date	1
character	2
numeric	4
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-04-12	2021-03-07	2020-09-23	95

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1	2	2	0	56	0
group	0	1	5	11	0	9	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
cases	15684	0.67	30240.68	103176.64	0	568	3661	21026	2619476
deaths	17686	0.63	708.93	1836.84	-1	12	68	440	24402
hosp	37253	0.22	2077.78	4654.37	0	67	345	1716	41099

tests	43549	0.09	349773.42	1269936.08	0	6298	36108	199214	18567612
-------	-------	------	-----------	------------	---	------	-------	--------	----------

The group variable is coded as follows:

```

groups
  White
  Black
  Latino
  Asian
  AI/AN
  NH/PI
  Multiracial
  Other
  Unknown

```

AI/AN is American Indian/Alaska Native. NH/PI is Native Hawaiian/Pacific Islander. State-level counts should be handled with care, given the widely varying population distribution of people of different racial backgrounds by state.

### Author(s)

Kieran Healy

### Source

<https://covidtracking.com/race>

---

<i>fmt_nc</i>	<i>fmt_nc</i>
---------------	---------------

---

### Description

Format *fmt\_nc* in *df*

### Usage

```
fmt_nc(x)
```

### Arguments

*x*                      *df*

### Details

use in `fn` documentation



**Value**

formatted string

**Author(s)**

Kieran Healy

**Examples**

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

*fmt\_nr*

*fmt\_nr*

---

**Description**

Format *fmt\_nr* in *df*

**Usage**

```
fmt_nr(x)
```

**Arguments**

*x*                    *df*

**Details**

use in fn documentation

**Value**

formatted string

**Author(s)**

Kieran Healy

**Examples**

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

MMWRweek2Date	<i>FUNCTION_TITLE</i>
---------------	-----------------------

---

**Description**

FUNCTION\_DESCRIPTION

**Usage**

MMWRweek2Date(MMWRyear, MMWRweek, MMWRday = NULL)

**Arguments**

MMWRyear	PARAM_DESCRIPTION
MMWRweek	PARAM_DESCRIPTION
MMWRday	PARAM_DESCRIPTION, Default: NULL

**Details**

DETAILS

**Value**

OUTPUT\_DESCRIPTION

**Author(s)**

Kieran Healy

**Source**<http://>**Examples**

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

MMWRweekday	<i>FUNCTION_TITLE</i>
-------------	-----------------------

---

**Description**

FUNCTION\_DESCRIPTION

**Usage**

MMWRweekday(date)

**Arguments**

date           PARAM\_DESCRIPTION

**Details**

DETAILS

**Value**

OUTPUT\_DESCRIPTION

**Author(s)**

Kieran Healy

**Source**<http://>**Examples**

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

mmwr\_week\_to\_date      *FUNCTION\_TITLE*

---

### Description

FUNCTION\_DESCRIPTION

### Usage

```
mmwr_week_to_date(year, week, day = NULL)
```

### Arguments

year	PARAM_DESCRIPTION
week	PARAM_DESCRIPTION
day	PARAM_DESCRIPTION, Default: NULL

### Details

DETAILS

### Value

OUTPUT\_DESCRIPTION

### Author(s)

Kieran Healy

### Source

<http://>

### See Also

[MMWRweek2Date](#)

### Examples

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

nchs\_sas

*Provisional COVID-19 Death Counts by Sex, Age, and State*


---

## Description

Deaths involving coronavirus disease (COVID-19), pneumonia, and influenza reported to NCHS by sex and age group and state.

## Usage

nchs\_sas

## Format

A tibble with 115,668 rows and 15 variables:

`data_as_of` date Date of data release

`start_date` date First date of data period

`end_date` date Last date of data period

`group` character Unit of time observation: whether data in this row are measured By month, By total, or By year

`year` integer Year of observation

`month` integer Month of observation

`state` character Jurisdiction of occurrence. One of: United States total, a US State, District of Columbia, *and New York City*, separate from New York state.

`sex` character Sex

`age_group` character Age group

`covid_19_deaths` integer Deaths involving COVID-19 (ICD-code U07.1)

`total_deaths` integer Deaths from all causes of death

`pneumonia_deaths` integer Pneumonia Deaths (ICD-10 codes J12.0-J18.9)

`pneumonia_and_covid_19_deaths` integer Deaths with Pneumonia and COVID-19 (ICD-10 codes J12.0-J18.9 and U07.1)

`influenza_deaths` integer Influenza Deaths (ICD-10 codes J09-J11)

`pneumonia_influenza_or_covid_19_deaths` integer Deaths with Pneumonia, Influenza, or COVID-19 (ICD-10 codes U07.1 or J09-J18.9)

## Details

Table: Data summary

Name	nchs_sas
Number of rows	115668

Number of columns	15
<hr/>	
Column type frequency:	
Date	1
character	6
numeric	8
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2023-01-18	2023-01-18	2023-01-18	1

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
start_date	0	1	10	10	0	37	0
end_date	0	1	10	10	0	37	0
group	0	1	7	8	0	3	0
state	0	1	4	20	0	54	0
sex	0	1	4	9	0	3	0
age_group	0	1	8	17	0	17	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
year	2754	0.98	2021.10	0.91	2020	2020	2020	2020	2020
month	13770	0.88	6.35	3.52	1	3	3	3	3
covid_19_deaths	31823	0.72	351.76	6263.51	0	0	0	0	0
total_deaths	17146	0.85	2812.18	52269.95	0	41	41	41	41
pneumonia_deaths	36293	0.69	349.71	6016.66	0	0	0	0	0
pneumonia_and_covid_19_deaths	30476	0.74	174.88	3162.39	0	0	0	0	0
influenza_deaths	22407	0.81	4.94	103.26	0	0	0	0	0
pneumonia_influenza_or_covid_19_deaths	35678	0.69	535.21	9239.91	0	0	0	0	0

Number of deaths reported in this table are the total number of deaths received and coded as of the date of analysis, and do not represent all deaths that occurred in that period. Data during this period are incomplete because of the lag in time between when the death occurred and when the death certificate is completed, submitted to NCHS and processed for reporting purposes. This delay can range from 1 week to 8 weeks or more. Missing values may indicate that a category has between 1 and 9 observed cases and have been suppressed in accordance with NHCS confidentiality standards. As of September 2, 2020, this data file

includes the following age groups in addition to the age groups that are routinely included: 0-17, 18-29, 30-49, and 50-64. The new age groups are consistent with categories used across CDC COVID-19 surveillance pages. When analyzing the file, the user should make sure to select only the desired age groups. Summing across all age categories provided will result in double counting deaths from certain age groups. Similarly, the state variable includes the United States as a whole, and New York City counted separately from the rest of New York State. The temporal unit of observation also varies, with totals given by year, by month, and overall. It is necessary to first filter the data by desired time unit, region, and age group to ensure there is no double-counting in subsequent calculations.

**Author(s)**

Kieran Healy

**Source**

National Center for Health Statistics [https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku](https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-State)

**References**

<https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku>

---

nchs\_wdc

*Weekly Counts of Deaths by State and Select Causes 2014-2021*

---

**Description**

Final counts of deaths by the week the deaths occurred, by state of occurrence, and by select causes of death for 2014-2018, and Provisional counts of deaths by the week the deaths occurred, by state of occurrence, and by select underlying causes of death for 2019-2020. The dataset also includes weekly provisional counts of death for COVID-19, coded to ICD-10 code U07.1 as an underlying or multiple cause of death.

**Usage**

nchs\_wdc

**Format**

A data frame with 347,706 rows and 7 variables:

`jurisdiction` character Jurisdiction of Occurrence

`year` double MMWR Year

`week` double MMWR Week

`week_ending_date` double MMWR Week ending date

`cause_detailed` character Cause with ICD Codes  
`n` double Count of deaths  
`cause` character Cause of death

## Details

For 2014-2019, death counts in this dataset were derived from the National Vital Statistics System database that provides the most timely access to the data. Therefore, counts may differ slightly from final data due to differences in processing, recoding, and imputation. For 2019-2021, the dataset also includes weekly provisional counts of death for COVID-19, coded to ICD-10 code U07.1 as an underlying or multiple cause of death. Number of deaths reported in this table are the total number of deaths received and coded as of the date of analysis, and do not represent all deaths that occurred in that period. Data for 2020 and 2021 are provisional and may be incomplete because of the lag in time between when the death occurred and when the death certificate is completed, submitted to NCHS and processed for reporting purposes. Causes of death included in this dataset are tabulated by underlying cause of death ICD-10 codes. COVID-19 deaths by underlying cause and multiple cause of death are also included.

## Author(s)

Kieran Healy

## Source

2014-2019: <https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-State-and-Select-Causes/3yf8-kanr>. 2020-2021: <https://data.cdc.gov/NCHS/Weekly-Counts-of-Deaths-by-State-and-Select-Causes/muzy-jte6>

---

nchs\_wss

*Provisional Death Counts for Coronavirus Disease (COVID-19):  
Weekly State-Specific Data Updates*

---

## Description

This report provides a weekly summary of deaths with coronavirus disease 2019 (COVID-19) by select geographic and demographic variables. In this release, counts of deaths are provided by the race and Hispanic origin of the decedent.

## Usage

nchs\_wss



**Format**

A tibble with 15,582 rows and 12 variables:

**data\_as\_of** date Date of analysis

**start\_date** date Start date of coverage

**end\_date** date End date of coverage

**year** character Year. One of "2020", "2021", or "2020/2021".

**month** dbl Month

**obs\_unit** character Unit of observation. One of: By Total, By Year, By Month.

**state** character Geographical unit. One of: the United States, a U.S. State, the District of Columbia, or New York City. New York state measures *do not* include New York City

**race\_ethnicity** chr Race and ethnic group. One of: Non-Hispanic White, Non-Hispanic Black or African American, Non-Hispanic American Indian or Alaska Native, Non-Hispanic Asian, Non-Hispanic Native Hawaiian or Other Pacific Islander, Non Hispanic more than one race, Hispanic or Latino.

**deaths** integer Count of deaths

**dist\_pct** double Distribution of COVID-19 deaths (%): Deaths for each group as a percent of the total number of COVID-19 deaths reported.

**uw\_dist\_pop\_pct** double Unweighted distribution of population (%): Population of each group as a percent of the total population.

**wt\_dist\_pop\_pct** double Weighted distribution of population (%): Population of each group as percent of the total population after accounting for how the race and Hispanic origin population is distributed in relation to the geographic areas impacted by COVID-19.

**Details**

Table: Data summary

Name	nchs_wss
Number of rows	15582
Number of columns	12
<hr/>	
Column type frequency:	
Date	1
character	6
numeric	5
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
data_as_of	0	1	2023-01-18	2023-01-18	2023-01-18	1

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
start_date	0	1	10	10	0	37	0
end_date	0	1	10	10	0	37	0
year	0	1	4	9	0	5	0
obs_unit	0	1	7	8	0	3	0
state	0	1	4	20	0	53	0
race_ethnicity	0	1	18	54	0	7	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
month	1855	0.88	6.35	3.52	1	3.0	6.0	9.0	12.0	
deaths	4625	0.70	596.40	8680.87	0	0.0	14.0	100.0	718968.0	
dist_pct	4625	0.70	17.59	29.22	0	0.0	1.1	19.7	100.0	
uw_dist_pop_pct	0	1.00	14.28	23.57	0	0.9	3.1	12.7	92.7	
wt_dist_pop_pct	0	1.00	13.68	21.60	0	0.5	3.2	14.4	93.6	

The percent of deaths reported in this table are the total number of represent all deaths received and coded as of the date of analysis and do not represent all deaths that occurred in that period. Data are incomplete because of the lag in time between when the death occurred and when the death certificate is completed, submitted to NCHS and processed for reporting purposes. *This delay can range from 1 week to 8 weeks or more*, depending on the jurisdiction, age, and cause of death. Provisional counts reported here track approximately 1–2 weeks behind other published data sources on the number of COVID-19 deaths in the U.S. COVID-19 deaths are defined as having confirmed or presumed COVID-19, and are coded to ICD-10 code U07.1. Unweighted population percentages are based on the Single-Race Population Estimates from the U.S. Census Bureau, for the year 2018 (available from: <https://wonder.cdc.gov/single-race-population.html>). Weighted population percentages are computed by multiplying county-level population counts by the count of COVID deaths for each county, summing to the state-level, and then estimating the percent of the population within each racial and ethnic group. These weighted population distributions therefore more accurately reflect the geographic locations where COVID outbreaks are occurring. Jurisdictions are included in this table if more than 100 deaths were received and processed by NCHS as of the data of analysis.

Race and Hispanic-origin categories are based on the 1997 Office of Management and Budget (OMB) standards (1,2), allowing for the presentation of data by single race and Hispanic origin. These race and Hispanic-origin groups—non-Hispanic single-race white, non-Hispanic single-race black or African American, non-Hispanic single-race American Indian

or Alaska Native (AIAN), non-Hispanic single-race Asian, and non-Hispanic single-race Native Hawaiian and Other Pacific Islander —differ from the bridged-race categories shown in most reports using mortality data.

New York State totals exclude New York City (provided in table separately).

Missing values may indicate that a category has between 1 and 9 observed cases and have been suppressed in accordance with NHCS confidentiality standards.

### Author(s)

Kieran Healy

### Source

National Center for Health Statistics <https://data.cdc.gov/NCHS/Provisional-Death-Counts-for-Coronavirus-pj7m-y5uh>

---

nssp_covid_er_nat	<i>NSSP National COVID-related ER Visits</i>
-------------------	--

---

### Description

National Syndromic Surveillance Program (NSSP): Emergency Department Visits and Percentage of Visits for COVID-19-Like Illness (CLI) or Influenza-like Illness (ILI)

### Usage

nssp\_covid\_er\_nat

### Format

A data frame with 54 rows and 9 variables:

```

week integer COLUMN_DESCRIPTION
num_fac integer COLUMN_DESCRIPTION
total_ed_visits character COLUMN_DESCRIPTION
visits integer COLUMN_DESCRIPTION
pct_visits double COLUMN_DESCRIPTION
visit_type character COLUMN_DESCRIPTION
region character COLUMN_DESCRIPTION
source character COLUMN_DESCRIPTION
year integer COLUMN_DESCRIPTION

```

**Details**

Table: Data summary

Name	nssp_covid_er_nat
Number of rows	54
Number of columns	9
<hr/>	
Column type frequency:	
character	4
numeric	5
<hr/>	
Group variables	None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
total_ed_visits	0	1	7	7	0	27	0
visit_type	0	1	3	3	0	2	0
region	0	1	8	8	0	1	0
source	0	1	21	21	0	1	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
week	0	1	26.04	19.81	1.00	7.25	14.00	45.75
num_fac	0	1	3346.89	48.97	3249.00	3329.50	3352.00	3389.50
visits	0	1	41521.67	16344.25	17639.00	31216.00	39183.50	50532.00
pct_visits	0	1	0.02	0.01	0.01	0.01	0.02	0.02
year	0	1	2019.52	0.50	2019.00	2019.00	2020.00	2020.00

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdcovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)).

**Author(s)**

Kieran Healy

**Source**

Courtesy of Bob Rudis's `cdcovidview` package

**References**

<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/04102020/nssp-regions.html>

---

nssp\_covid\_er\_reg      *NSSP Regional COVID ER Visits*

---

**Description**

Regional Syndromic Surveillance Program (NSSP): Emergency Department Visits and Percentage of Visits for COVID-19-Like Illness (CLI) or Influenza-like Illness (ILI)

**Usage**

nssp\_covid\_er\_reg

**Format**

A tibble with 538 rows and 9 variables:

```

week integer COLUMN_DESCRIPTION
num_fac integer COLUMN_DESCRIPTION
total_ed_visits character COLUMN_DESCRIPTION
visits integer COLUMN_DESCRIPTION
pct_visits double COLUMN_DESCRIPTION
visit_type character COLUMN_DESCRIPTION
region character COLUMN_DESCRIPTION
source character COLUMN_DESCRIPTION
year integer COLUMN_DESCRIPTION

```

**Details**

Table: Data summary

Name	nssp_covid_er_reg
Number of rows	538
Number of columns	9
<hr/>	
Column type frequency:	
character	4
numeric	5
<hr/>	
Group variables	None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
total_ed_visits	0	1	5	6	0	269	0
visit_type	0	1	3	3	0	2	0
region	0	1	8	9	0	10	0
source	0	1	21	21	0	1	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
week	0	1	25.99	19.66	1	7.00	14.00	46.00	52.00
num_fac	0	1	335.18	234.58	135	190.00	222.00	343.00	884.00
visits	0	1	4164.87	4028.53	279	1596.00	2780.00	4723.75	23345.00
pct_visits	0	1	0.02	0.01	0	0.01	0.02	0.02	0.11
year	0	1	2019.52	0.50	2019	2019.00	2020.00	2020.00	2020.00

The U.S. Centers for Disease Control provides weekly summary and interpretation of key indicators that have been adapted to track the COVID-19 pandemic in the United States. Data is retrieved using the `cdccovidview` package from both COVIDView (<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>) and COVID-NET ([https://gis.cdc.gov/grasp/COVIDNet/COVID19\\_3.html](https://gis.cdc.gov/grasp/COVIDNet/COVID19_3.html)).

**Author(s)**

Kieran Healy

**Source**

Courtesy of Bob Rudis's `cdccovidview` package

**References**

<https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/04102020/nssp-regions.html>

---

nytcovcounty

*NYT COVID-19 data for US counties, current as of Sunday, January 22, 2023*

---

**Description**

A dataset containing US county-level data on COVID-19, collected by the New York Times.

**Usage**

nytcovcounty

**Format**

A tibble with 2,502,832 rows and 6 columns

**date** Date in YYYY-MM-DD format (date)

**county** County name (character)

**state** State name (character)

**fips** County FIPS code (character)

**cases** Cumulative N reported cases

**deaths** Cumulative N reported deaths

**Details**

Table: Data summary

Name	nytcovcounty
Number of rows	2502832
Number of columns	6
<hr/>	
Column type frequency:	
Date	1
character	3
numeric	2
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-01-21	2022-05-13	2021-04-23	844

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
county	0	1.00	3	35	0	1932	0
state	0	1.00	4	24	0	56	0
fips	23678	0.99	5	5	0	3220	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
cases	0	1.00	10033.80	47525.22	0	382	1773	5884	2908425	
deaths	57605	0.98	161.61	820.33	0	6	33	101	40267	

## Source

The New York Times <https://github.com/nytimes/covid-19-data> For details on the methods and limitations see <https://github.com/nytimes/covid-19-data>. For county data, note in particular:

- New York: All cases for the five boroughs of New York City (New York, Kings, Queens, Bronx and Richmond counties) are assigned to a single area called New York City. There is a large jump in the number of deaths on April 6th due to switching from data from New York City to data from New York state for deaths. For all New York state counties, starting on April 8th we are reporting deaths by place of fatality instead of residence of individual.
- Kansas City, Mo: Four counties (Cass, Clay, Jackson and Platte) overlap the municipality of Kansas City, Mo. The cases and deaths that we show for these four counties are only for the portions exclusive of Kansas City. Cases and deaths for Kansas City are reported as their own line.
- Alameda County, Calif: Counts for Alameda County include cases and deaths from Berkeley and the Grand Princess cruise ship.
- Douglas County, Neb. Counts for Douglas County include cases brought to the state from the Diamond Princess cruise ship.
- Chicago: All cases and deaths for Chicago are reported as part of Cook County.
- Guam: Counts for Guam include cases reported from the USS Theodore Roosevelt.

---

<code>nytcovstate</code>	<i>NYT COVID-19 data for the US states, current as of Sunday, January 22, 2023</i>
--------------------------	--

---

## Description

A dataset containing US state-level data on COVID-19, collected by the New York Times.

## Usage

```
nytcovstate
```

## Format

A tibble with 58,526 rows and 5 columns

**date** Date in YYYY-MM-DD format (date)

**state** State name (character)

**fips** State FIPS code (character)

**cases** Cumulative N reported cases

**deaths** Cumulative N reported deaths



**Details**

Table: Data summary

Name	nytcovstate
Number of rows	58526
Number of columns	5
<hr/>	
Column type frequency:	
Date	1
character	2
numeric	2
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-01-21	2023-01-21	2021-08-16	1097

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1	4	24	0	56	0
fips	0	1	2	2	0	56	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
cases	0	1	834511.91	1394631.70	1	64160	324958	985279.8	1195560
deaths	0	1	11294.84	16797.98	0	1080	4790	14373.0	10198

**Source**

The New York Times <https://github.com/nytimes/covid-19-data>. For details on the methods and limitations see <https://github.com/nytimes/covid-19-data>.

---

nytcovus	<i>NYT COVID-19 data for the US, current as of Sunday, January 22, 2023</i>
----------	---

---

### Description

A dataset containing US national-level data on COVID-19, collected by the New York Times.

### Usage

```
nytcovus
```

### Format

A tibble with 1,097 rows and 3 columns

**date** Date in YYYY-MM-DD format (date)

**cases** Cumulative N reported cases

**deaths** Cumulative N reported deaths

### Details

Table: Data summary

Name	nytcovus
Number of rows	1097
Number of columns	3
<hr/>	
Column type frequency:	
Date	1
numeric	2
<hr/>	
Group variables	None

#### Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	0	1	2020-01-21	2023-01-21	2021-07-22	1097

#### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
---------------	-----------	---------------	------	----	----	-----	-----	-----

cases	0	1	44522009.0	35239239.4	1	8404635	34364829	80836264	10
deaths	0	1	602590.7	370532.5	0	222195	609870	989584	

## Source

The New York Times <https://github.com/nytimes/covid-19-data>. For details on the methods and limitations see <https://github.com/nytimes/covid-19-data>.

---

<b>nytexcess</b>	<i>NYT Excess Mortality Estimates, current as of Sunday, January 22, 2023</i>
------------------	---

---

## Description

All-cause mortality is widely used by demographers and other researchers to understand the full impact of deadly events, including epidemics, wars and natural disasters. The totals in this data include deaths from Covid-19 as well as those from other causes, likely including people who could not be treated or did not seek treatment for other conditions.

## Usage

```
nytexcess
```

## Format

A tibble with 7,258 rows and 12 columns

**country** character Country Name

**placename** character Place Name

**frequency** character Reporting period. Weekly or monthly, depending on how the data is recorded.

**start\_date** date The first date included in the period.

**end\_date** date The last date included in the period,

**year** character Year of data. Note that this variable is of type character and not integer because several observations are notes to the effect that the year is an average of two years.

**month** integer Numerical month.

**week** integer Numerical week.

**deaths** integer The total number of confirmed deaths recorded from any cause.

**expected\_deaths** integer The baseline number of expected deaths, calculated from a historical average. See details below.

**excess\_deaths** integer The number of deaths minus the expected deaths.

**baseline** character The years used to calculate expected\_deaths.

**Details**

Table: Data summary

Name	nytexcess
Number of rows	7258
Number of columns	12
<hr/>	
Column type frequency:	
Date	2
character	5
numeric	5
<hr/>	
Group variables	None

**Variable type: Date**

skim_variable	n_missing	complete_rate	min	max	median	n_unique
start_date	768	0.89	2010-01-09	2020-12-23	2018-02-05	1267
end_date	768	0.89	2010-01-15	2020-12-29	2018-02-11	1267

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
country	0	1.00	4	14	0	35	0
placename	6883	0.05	6	8	0	4	0
frequency	0	1.00	6	7	0	2	0
year	0	1.00	4	17	0	15	0
baseline	5990	0.17	20	25	0	7	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
month	0	1.00	6.60	3.36	1	4.00	7.0	9.0	12.0
week	666	0.91	26.77	14.58	2	14.00	27.0	39.0	52.0
deaths	0	1.00	7968.24	14334.14	455	1460.00	2395.5	10486.0	14129.0
expected_deaths	5990	0.17	9237.09	15850.00	548	1443.00	2423.0	10771.5	13934.0
excess_deaths	5990	0.17	1195.43	3242.72	-6721	-42.25	76.5	926.0	3040.0

Expected deaths for each area based on historical data for the same time of year. These expected deaths are the basis for our excess death calculations, which estimate how many more people have died this year than in an average year.

The number of years used in the historical averages changes depending on what data is available, whether it is reliable and underlying demographic changes. See Data Sources for the years used to calculate the baselines. The baselines do not adjust for changes in age or other demographics, and they do not account for changes in total population.

The number of expected deaths are not adjusted for how non-Covid-19 deaths may change during the outbreak, which will take some time to figure out. As countries impose control measures, deaths from causes like road accidents and homicides may decline. And people who die from Covid-19 cannot die later from other causes, which may reduce other causes of death. Both of these factors, if they play a role, would lead these baselines to understate, rather than overstate, the number of excess deaths.

### Author(s)

Kieran Healy

### Source

The New York Times <https://github.com/nytimes/covid-19-data/tree/master/excess-deaths>.

### References

For further details on these data see <https://github.com/nytimes/covid-19-data/tree/master/excess-deaths>

---

start_date	FUNCTION_TITLE
------------	----------------

---

### Description

FUNCTION\_DESCRIPTION

### Usage

```
start_date(year)
```

### Arguments

year	PARAM_DESCRIPTION
------	-------------------

### Details

DETAILS

### Value

OUTPUT\_DESCRIPTION

**Author(s)**

AUTHOR\_NAME

**Source**

http://

**Examples**

```
## Not run:
if(interactive()){
  #EXAMPLE1
}

## End(Not run)
```

---

stmf

---

*Short Term Mortality Fluctuations (STMF) data series*


---

**Description**

Human Mortality Database (HMD) series of weekly death counts across countries.

**Usage**

stmf

**Format**

A tibble with 580,395 rows and 17 variables:

**country\_code** Mortality database country code

**cname** character Country name

**iso2** character ISO2 country code

**iso3** character ISO3 country code

**year** double Year

**week** double Week number. Each year in the STMF refers to 52 weeks, each week has 7 days. In some cases, the first week of a year may include several days from the previous year or the last week of a year may include days (and, respectively, deaths) of the next year. In particular, it means that a statistical year in the STMF is equal to the statistical year in annual country-specific statistics.

**sex** character Sex. m = Males. f = Females. b = Both combined.

**split** double Indicates if data were split from aggregated age groups (0 if the original data has necessary detailed age scale). For example, if the original age scale was 0-4, 5-29, 30-65, 65+, then split will be equal to 1



Norway	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Poland	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Portugal	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Russian Federation	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Scotland	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Slovakia	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Slovenia	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Spain	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Sweden	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Switzerland	-	-	-	-	-	-	-	-	-	-	-	Y	Y
Taiwan, Province of China	-	-	-	-	-	-	-	-	-	-	-	Y	Y
United States	-	-	-	-	-	-	-	-	-	-	-	-	-

## Variables Table: Data summary

Name	stmf
Number of rows	580395
Number of columns	17
<hr/>	
Column type frequency:	
Date	1
character	7
numeric	9
<hr/>	
Group variables	None

## Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
approx_date	0	1	1990-01-07	2023-01-01	2012-10-07	1722

## Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
country_code	0	1.00	3	7	0	38	0
cname	0	1.00	5	25	0	38	0
iso2	34380	0.94	2	2	0	35	0
continent	35850	0.94	4	13	0	5	0
iso3	34380	0.94	3	3	0	35	0
sex	0	1.00	1	1	0	3	0
age_group	0	1.00	3	5	0	5	0

## Variable type: numeric



skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
year	0	1	2011.58	6.88	1990	2006.00	2012.00	2017.00	2022.00
week	0	1	26.50	15.03	1	13.00	26.00	39.00	53.00
split	0	1	0.12	0.32	0	0.00	0.00	0.00	1.00
split_sex	0	1	0.00	0.07	0	0.00	0.00	0.00	1.00
forecast	0	1	0.10	0.30	0	0.00	0.00	0.00	1.00
death_count	0	1	617.60	1585.49	0	39.00	162.00	449.75	26362.00
death_rate	0	1	0.05	0.07	0	0.00	0.02	0.07	0.57
deaths_total	0	1	3088.00	6498.29	2	472.00	998.00	2543.00	87413.00
rate_total	0	1	0.01	0.00	0	0.01	0.01	0.01	0.04

**Author(s)**

Kieran Healy

**Source**Human Mortality Database, <http://mortality.org>**References**

"Short-term Mortality Fluctuations Dataseries" n.d., [https://www.mortality.org/Public/STMF\\_DOC/STMFNote.pdf](https://www.mortality.org/Public/STMF_DOC/STMFNote.pdf)

---

`stmf_country_years`    *Make a table of stmf country years*

---

**Description**

Make a table of stmf country years

**Usage**

```
stmf_country_years(df = stmf)
```

**Arguments**

`df`                    The stmf data frame

**Details**

Get a table of country x year coverage for stmf

**Value**

A tibble

**Author(s)**

Kieran Healy

**Source**<http://>**Examples**

```
## Not run:  
if(interactive()){  
  #EXAMPLE1  
}  
  
## End(Not run)
```

---

`tabular`*tabular*

---

**Description**

Make an Rd table from a data frame

**Usage**`tabular(df, ...)`**Arguments**

<code>df</code>	Data frame
<code>...</code>	Other args

**Details**

DETAILS

**Value**

Rd table

**Author(s)**

Kieran Healy

**Source**<http://>

**Examples**

```
## Not run:
if(interactive()){
  #EXAMPLE1
}

## End(Not run)
```

---

 uspop

*State population estimates for US States*


---

**Description**

Population estimates for US States as of July 1st 2018

**Usage**

```
uspop
```

**Format**

A tibble with 459 rows and 17 variables:

```
state character State Name
state_abbr character State Abbreviation
statefips character 2-digit FIPS code
region_name character Census region
division_name character Census Division
sex_id character Sex id
sex character Sex label
hisp_id character Ethnicity: Hispanic id
hisp_label character Hispanic label
fips character Full FIPS code
pop double Total population
white double Race alone: White
black double Race alone: Black or African-American
amind double Race alone: American Indian and Alaska Native
asian double Race alone: Asian
nhopi double Race alone: Native Hawaiian and Other Pacific Islander
tom double Race alone: Two or more races
```

**Details**

Table: Data summary

Name	uspop
Number of rows	459
Number of columns	17
<hr/>	
Column type frequency:	
character	10
numeric	7
<hr/>	
Group variables	None

**Variable type: character**

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
state	0	1.00	4	20	0	51	0
state_abbr	9	0.98	2	2	0	50	0
statefips	0	1.00	2	2	0	51	0
region_name	9	0.98	4	9	0	4	0
division_name	9	0.98	7	18	0	9	0
sex_id	0	1.00	4	6	0	3	0
sex	0	1.00	4	10	0	3	0
hisp_id	0	1.00	4	7	0	3	0
hisp_label	0	1.00	5	12	0	3	0
fips	0	1.00	11	11	0	51	0

**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
pop	0	1	2851132.32	4198641.26	6154	386961.5	1349442	3558480.0
white	0	1	2179861.40	3116129.25	5120	296294.0	1088503	2759335.5
black	0	1	381736.98	644380.66	260	11907.0	80714	486281.5
amind	0	1	36143.97	65036.83	161	6103.5	15273	35770.5
asian	0	1	168458.39	515557.14	79	5045.5	26484	140424.5
nhopi	0	1	6966.61	18657.18	23	669.0	2029	5063.5
tom	0	1	77964.97	131251.16	455	12091.0	33757	98669.5

U.S. Census estimates. Be aware of the US Census classifications of Race and Ethnicity. For the estimated total population for each State, jointly filter on `totsex` in `sex_id` and `tothisp` in `hisp_id` and then select `pop`.

**Author(s)**

Kieran Healy

**Source**

<https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-detail.html>

**References**

<https://www2.census.gov/programs-surveys/popest/tables/2010-2018/state/asrh/PEPSR6H.pdf>

---

%nin%

%nin%

---

**Description**

Convenience 'not-in' operator

**Usage**

`x %nin% y`

**Arguments**

`x`                vector of items  
`y`                vector of all values

**Details**

Complement of the built-in operator `%in%`. Returns the elements of `x` that are not in `y`.

**Value**

logical vector of items in `x` not in `y`

**Author(s)**

Kieran Healy

**Examples**

```
fruit <- c("apples", "oranges", "banana")
"apples" %nin% fruit
"pears" %nin% fruit
```

# Index

## \* datasets

apple\_mobility, 2  
cdc\_catchments, 4  
cdc\_deaths\_by\_age, 6  
cdc\_deaths\_by\_sex, 8  
cdc\_deaths\_by\_state, 10  
cdc\_deaths\_by\_week, 12  
countries, 14  
covnat\_daily, 15  
covnat\_weekly, 16  
covus, 18  
covus\_ethnicity, 21  
covus\_race, 22  
nchs\_sas, 29  
nchs\_wdc, 31  
nchs\_wss, 32  
nssp\_covid\_er\_nat, 35  
nssp\_covid\_er\_reg, 37  
nytcovcounty, 38  
nytcovstate, 40  
nytcovus, 42  
nytexcess, 43  
stmf, 46  
uspop, 51  
%nin%, 53

apple\_mobility, 2

cdc\_catchments, 4  
cdc\_deaths\_by\_age, 6  
cdc\_deaths\_by\_sex, 8  
cdc\_deaths\_by\_state, 10  
cdc\_deaths\_by\_week, 12  
countries, 14  
covnat\_daily, 15  
covnat\_weekly, 16  
covus, 18  
covus\_ethnicity, 21  
covus\_race, 22

fmt\_nc, 24  
fmt\_nr, 25  
mmwr\_week\_to\_date, 28  
MMWRweek2Date, 26, 28  
MMWRweekday, 27  
nchs\_sas, 29  
nchs\_wdc, 31  
nchs\_wss, 32  
nssp\_covid\_er\_nat, 35  
nssp\_covid\_er\_reg, 37  
nytcovcounty, 38  
nytcovstate, 40  
nytcovus, 42  
nytexcess, 43  
start\_date, 45  
stmf, 46  
stmf\_country\_years, 49  
tabular, 50  
uspop, 51