

Package ‘troopdata’

October 7, 2026

Title Tools for Analyzing Cross-National Military Deployment and Basing Data

Version 1.1.0

Maintainer Michael Flynn <meflynn@ksu.edu>

Description These functions generate data frames on troop deployments and military basing using U.S. Department of Defense data on overseas military deployments. This package provides functions for pulling country-year troop deployment and basing data. Subsequent versions will hopefully include cross-national data on deploying countries.

License GPL (>= 3)

Language en-US

Encoding UTF-8

LazyData true

URL <https://github.com/meflynn/troopdata>,
<https://meflynn.github.io/troopdata/>

BugReports <https://github.com/meflynn/troopdata/issues>

Depends R (>= 4.1.0)

Imports magrittr, dplyr, rlang

Suggests testthat (>= 3.0.0), furrr, tidyselect, spelling, knitr,
rmarkdown, viridis, ggplot2, countrycode, tidyr, stringr,
haven, maps, pdftools

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Michael Flynn [aut, cre] (ORCID:
<<https://orcid.org/0000-0002-6064-740X>>)

Repository CRAN

Date/Publication 2026-10-07 08:30:02 UTC

Contents

troopdata-package	2
basedata	3
builddata	4
build_data_20260918	4
get_basedata	7
get_builddata	9
get_exercises	13
get_troopdata	15
mme_long	21
troopdata_rebuild_long	22
troopdata_rebuild_reports	26
troopdata_rebuild_us_states	30
Index	32

troopdata-package	<i>Tools for Analyzing Cross-National Military Deployment and Basing Data</i>
-------------------	---

Description

Generate customized data frames containing U.S. military deployment data

Details

The troopdata package contains functions for generating customized data sets on U.S. military deployment data and U.S. overseas basing data. The data contain country identifiers that can be used to merge these data with other data to analyze trends in the United States’ overseas military posture.

Author(s)

Michael E. Flynn <meflynn@ksu.edu>

References

<https://www.m-flynn.com/> <https://github.com/meflynn/troopdata>

See Also

Useful links:

- <https://github.com/meflynn/troopdata>
- <https://meflynn.github.io/troopdata/>
- Report bugs at <https://github.com/meflynn/troopdata/issues>

basedata	<i>Vine's U.S. basing data</i>
----------	--------------------------------

Description

basedata returns a data frame containing David Vine's US basing data.

Usage

basedata

Format

A data frame with country-base observations including the following variables:

countryname A character vector of country names.

ccode A numeric vector of Correlates of War country codes.

iso3c A character vector of ISO three character country codes.

basename Name of the facility.

lat The facility's latitude.

lon The facility's longitude.

base Binary indicator identifying the facility as a major base or not.

lilypad A binary indicator identifying the facility as a lilypad or not. Vine codes lilypads as less than 200 personnel or "other site" designation in Pentagon reports.

fundedstate A binary variable indicating whether or not the facility is a host-state base funded by the US.

Value

Returns the full data frame containing country observations of US military bases from the Cold War period through 2018.

Source

https://aura.american.edu/articles/online_resource/Lists_of_U_S_Military_Bases_Abroad_1776-2020/23856486

builddata	<i>U.S. Military overseas construction spending data (superseded)</i>
-----------	---

Description

builddata is the original geocoded construction spending data, compiled from operations and maintenance records rather than the Comptroller Annual Report C-1 exhibits. It is retained for reproducibility of earlier analyses. `get_builddata()` reads `build_data_20260918` instead.

Usage

```
builddata
```

Format

A data frame with location-year observations including the following variables:

countryname A character vector of country names.

cocode A numeric vector of Correlates of War country codes.

year Year of observed country-year spending.

iso3c A character vector of ISO three character country codes.

location Name of the facility where spending occurred, or host country where detailed facility information is unavailable.

spend_construction Total obligational authority associated with the observed location-year in thousands of current US dollars.

lat The facility's latitude.

lon The facility's longitude.

Value

Returns a data frame containing location-year observations of U.S. military construction spending data from 2008-2019.

build_data_20260918	<i>U.S. Military construction spending data</i>
---------------------	---

Description

build_data_20260918 returns a data frame containing geocoded location-project-year military construction spending data, for projects in the United States and overseas.

Usage

```
build_data_20260918
```

Format

A data frame of location-project-year observations including the following variables:

`fiscal_year` Fiscal year of the project.

`state_country` Two-letter state or country code as reported. Overseas codes mix ISO2C with legacy FIPS 10-4 and DoD codes.

`state_country_name` Readable form of the state or country as reported. The only form available for codes such as ZU (Unspecified Worldwide Locations), which have no country identifier.

`iso3c` ISO three character country code. NA for unspecified and worldwide locations. The 392 lines of FY2007-FY2016 that are filed under "Unspecified Worldwide Locations" but name a U.S. installation in their project title ("USA-224: Fort Hood, TX") carry "USA"; `state_country` and `location_name` are left as the report prints them, and the installation and state are in `location_full_name`.

`gwcode` Gleditsch and Ward country code. The G&W list has no code for a dependency, so territories carry the custom codes the troop data use for the same places: Puerto Rico 6, Greenland 1002, Diego Garcia 1004, Guam 1008, the Northern Mariana Islands 1011, the U.S. Virgin Islands 1013, Wake Island 1014, American Samoa 1041 and Ascension Island 1042. NA for unspecified and worldwide locations.

`location_name` Installation or location, in title case with initialisms preserved. The reports write the same place in block capitals one year and title case the next; a derived list of protected tokens keeps forms like "MCAS Iwakuni", "MacDill AFB" and "RAF Lakenheath" intact.

`location_name_reported`, `project_title_reported` The location and project title as the source sheet wrote them, before case harmonization.

`location_full_name` The address sent to the geocoder: the location followed by the country, or by the state and "United States". NA for rows that are not geocoded.

`location_code` Comptroller location code.

`latitude`, `longitude` Coordinates of the location, kept only where the geocoded point falls inside the country or state the row is filed under. NA for rows filed under no country (unspecified, worldwide and classified locations), for locations that are a label rather than a place ("Various Locations"), and where no point inside the country or state was found. A point inside the right country is not always the installation: a geocoder that does not know a place answers with the centre of the country, or with another place of the same name. The installations found to be placed that way are corrected by hand (`geo_source` is "manual"); the few that could not be located have no coordinates.

`geo_source` Where the coordinates came from: the geocoding service that resolved the location ("arcgis" for every geocoded row in the current data), or "manual" for a point entered by hand from `data-raw/geocode_overrides.csv`, which records what each point is and the page it was read from.

`organization` Service or agency, harmonized across years to full names: "Army", "Navy", "Air Force", "Defense-Wide", "Special Operations Command" and other defense agencies. Marine Corps construction is carried in the Navy accounts and has no value of its own; guard and reserve components appear in `appn_title`.

`organization_reported` The organization string as the source sheet wrote it, before harmonization.

`project_number, project_title` Project identifier and description. The title is in title case with initialisms preserved.

`appn_amount, auth_amount, auth_appn_amount, toa_amount` Appropriation, authorization, authorization of appropriation, and total obligational authority, in thousands of current US dollars. Negative values occur for rescissions and reprogramming.

`appn_title` Appropriation account title.

`budget_activity` Budget activity number. Meaningful only within an appropriation account: activity 1 is major construction in the MILCON accounts but FY 2005 BRAC, new construction or direct loan subsidy in others.

`budget_activity_title` Budget activity, harmonized to one label per activity across years. The source workbooks write these inconsistently, with case variants, abbreviations, truncations and typos.

`budget_activity_title_reported` The budget activity label as the source sheet wrote it, before harmonization.

`facility_category_title_reported` The facility title as the source sheet wrote it, before case and abbreviation harmonization.

`facility_category_code, facility_category_title` The three-digit DoD facility category code and its title, in sentence case with initialisms preserved. One code maps to exactly one title.

`facility_group_title` The coarse level of the DoD facility taxonomy, in sentence case. The reports carry it under three different headers – "Facility Group Title", "Fiscal Category Title", and, in the most recent years, the facility category column – and all three are collected here. Kept separate from `facility_category_title`, the finer three-digit level of the same taxonomy, though the split is a reconstruction rather than something the reports maintain; `get_builddata()` therefore searches both columns from its single `facility_category` argument.

`transaction_type` Type of budget action, where reported ("BUDGET", "REPROGRAM", rescissions). NA for most rows.

`sheet_type` Which kind of sheet the row came from: "base" for regular fiscal-year sheets, or "oco" for an overseas contingency operations exhibit. The other supplemental exhibits restate lines of the regular sheets and are not in the data.

`source_file, source_sheet, source_row` Workbook and sheet the row was read from, and its position in that sheet counting from 1 below the header. Together they identify one line of one report. A report can list two projects that agree in every other column, and this is what tells them apart.

`report_year` Latest fiscal year covered by the source workbook, i.e. the report vintage. Consecutive reports restate the same fiscal years, and every fiscal year here comes from the most recent report that covers it.

`is_request` TRUE where the row is a budget request rather than an enacted figure: the fiscal year is the terminal year of its source report, or the row comes from a request sheet. That is all of FY2016, whose later reports are not among the source files, and 26 lines of FY2010 from the overseas contingency operations request sheet ("C1_2010_OCO_Req") that are on no enacted sheet.

`classification` Security classification marking carried on the row, "U" for unclassified. FY2011 onward.

dollar_type Kind of dollar figure the row reports ("Budget" or "Congressional Rescission"). FY2003-FY2005.

existing_mission Whether the project supports an existing mission, as reported. FY2003-FY2005.

pe Program element. FY2003-FY2004.

percent_recapitalization Share of the project classed as recapitalization. No values: only the FY2008 report printed it, and the years it covers are taken from later reports.

state_country_sort Sort key the source sheet used for the state or country column. FY2006.

Value

Returns the full data frame containing location-project-year observations of U.S. military construction spending, compiled from the DoD Comptroller Annual Report C-1 exhibits. All amounts are in thousands of current US dollars. Each fiscal year is taken from the most recent report that covers it, so a project-year appears once, and each row is one line of that report.

Year coverage

The C-1 exhibit layout changed over time, and a later report does not always print what an earlier one did, so several fields are reported only for a subset of years: transaction_type and appn_title FY2001-FY2002; pe FY2003-FY2004; dollar_type and existing_mission FY2003-FY2005; location_code FY2000-FY2006; facility_category_code FY2000 and FY2003-FY2006; project_number FY2000-FY2006 and FY2013 onward; state_country_sort FY2006; classification FY2011 onward; facility_category_title FY2003-FY2005, FY2007-FY2021 and FY2026; facility_group_title FY2000 and FY2007 onward; state_country_name FY2006 and FY2009 onward. FY2006 reports appropriations only, so toa_amount, auth_amount, auth_appn_amount and budget_activity_title are missing for that year; toa_amount and auth_appn_amount are also missing for FY2000, and budget_activity_title for FY2000 and FY2008. The workbooks that supply FY2004-FY2007 give the service of an Army, Navy or Air Force line in their "Treasury Agency" column rather than under "Organization"; organization is filled from it, and organization_reported holds the letter printed there ("A", "N" or "F"). The series runs FY2000 through FY2026.

get_basedata

Function to retrieve customized U.S. basing data

Description

get_basedata() generates a customized data frame containing data obtained from David Vine's U.S. basing data.

Usage

```
get_basedata(host = NA, country_count = FALSE, groupvar = NULL)
```

Arguments

host	The Correlates of War (COW) numeric country code or ISO3C code for the host country or countries in the series
country_count	Logical. Should the function return a country-level count of the total number of bases or the country-site data
groupvar	A character string indicating how country count totals should be generated. Accepted values are 'countryname', 'ccode', or 'iso3c'. Can take on Required when using country_count argument.

Details

Our research team updated these data through 2018.

Value

`get_basedata()` returns a data frame containing information on U.S. military bases present within selected host countries. This can be customized to include country-base observations or country-count observations.

Author(s)

Michael E. Flynn

References

David Vine. 2015. Base Nation. Metropolitan Books. New York, NY.

Examples

```
## Not run:
library(tidyverse)
library(troopdata)

example <- get_basedata(host = NA)

head(example)

## End(Not run)
```

get_bulddata	<i>Function to retrieve customized U.S. military construction spending data.</i>
--------------	--

Description

get_bulddata() subsets build_data_20260918 to a customized data frame containing location-project-year observations of U.S. military construction and housing spending, as published in the U.S. Department of Defense Comptroller Annual Report C-1 exhibits. Users can filter by host country/location, organization, transaction type, facility title, budget activity, project title, spending type, and amount range.

Usage

```
get_bulddata(
  host = NA,
  location = NA,
  organization = NA,
  transaction_type = NA,
  facility_category = NA,
  budget_activity = NA,
  project = NA,
  spend_type = "all",
  min_amount = NA,
  max_amount = NA,
  include_requests = TRUE,
  startyear,
  endyear
)
```

Arguments

host	Character or numeric. If character, the three-letter ISO3C country code (iso3c field), or a vector of such codes (e.g., c("KWT", "JPN", "DEU") for Kuwait, Japan, and Germany). Matching is case-insensitive. If numeric, the value is treated as a Gleditsch and Ward (G&W) country code, or vector of codes (gwcode field), matched exactly. Domestic (US) construction is coded as "USA" / 2. Use NA (the default) to return all locations.
------	--

The G&W system does not assign codes to dependencies. Those locations carry the custom codes that the troop deployment data use for the same places, so one code selects a place in both data sets: Puerto Rico (6, "PRI"), Greenland (1002, "GRL"), Diego Garcia (1004, "IOT"), Guam (1008, "GUM"), the Northern Mariana Islands (1011, "MNP"), the U.S. Virgin Islands (1013, "VIR"), Wake Island (1014, "UMI"), American Samoa (1041, "ASM") and Ascension Island (1042, "SHN"). Roughly a quarter of rows report an amount but no usable location ("Unspecified Worldwide Locations" and similar) and have neither code.

location	Character. A string or regular expression to match against location_name and location_full_name. Matching is case-insensitive. A vector of patterns is collapsed into a single alternation. Use NA (the default) to return all locations.
organization	Character. One or more organizations to filter by, given as full names: "Army", "Navy", "Air Force", "Defense-Wide", "Special Operations Command", "Defense Logistics Agency", "Defense Health Agency", "Department of Defense Education Activity", "National Security Agency" and other defense agencies. Matching is case-insensitive and unmatched values raise an error. Use NA (the default) to include all organizations. See <code>sort(unique(troopdata::build_data_20260918\$organization))</code> for the full list, and <code>organization_reported</code> for the abbreviation each source sheet used. There is no "Marine Corps" value: the C-1 exhibits carry Marine Corps construction in the Navy accounts. Guard and reserve components appear in <code>appn_title</code> rather than here. For FY2004-FY2007 the service of an Army, Navy or Air Force line is read from the "Treasury Agency" column of the report, which is where those workbooks print it.
transaction_type	Character. One or more transaction types to include: "BUDGET", "REPROGRAM", "REDUCTION", "CONGRESSIONAL RESCISSION", or "PUBLIC LAW RESCISSION". Use NA (the default) to include all transaction types. NOTE: this field is reported for FY2001 and FY2002 only, so filtering on it restricts the result to those years.
facility_category	Character. One or more facility titles to filter by, at either level of the DoD facility taxonomy: the coarse group ("Operational facilities", "Family housing operations", "Training facilities") or the finer three-digit category ("Training buildings", "Airfield operational buildings", "Aircraft maintenance facilities"). Matching is case-insensitive and unmatched values raise an error. Use NA (the default) to include all facility titles. A value is matched against <code>facility_group_title</code> and <code>facility_category_title</code> together, and a row is kept if either column holds it. The two columns are levels of one vocabulary and the reports do not keep them apart: the coarse labels arrive under the header "Facility Group Title" in some years, "Fiscal Category Title" in others, and in the FY2022-forward workbooks in the facility category column itself. Separating them in the data is therefore a reconstruction, and searching both is the only way to filter without having to know which header a given report year used. Three labels ("Energy conservation", "Family housing P&D", "NATO security investment program") are genuinely present at both levels and return rows from each. See <code>sort(unique(c(troopdata::build_data_20260918\$facility_group_title, troopdata::build_data_20260918\$facility_category_title)))</code> for the 190 values present in the data. Not every row carries a facility title, so filtering on this argument drops the rest.
budget_activity	Character. One or more budget activity titles to filter by (e.g., "MAJOR CONSTRUCTION", "PLANNING AND DESIGN", "LEASING"). These labels are stored in upper case, but matching is case-insensitive. Use NA (the default) to include all budget activities. Reported for FY2001 onward.

project	Character. A string or regular expression to match against project_title. Matching is case-insensitive. A vector of patterns is collapsed into a single alternation. Use NA (the default) to return all projects.
spend_type	Character. The amount column used for dropping missing values and for applying min_amount / max_amount. One of: • "all" (default): no amount column is required to be present; amount filters are applied to toa_amount • "appn": appropriations (appn_amount), FY2000 onward • "auth": authorizations (auth_amount), FY2000 onward • "auth_appn": authorized for appropriation (auth_appn_amount), FY2001 onward • "toa": total obligational authority (toa_amount), FY2001 onward When spend_type is not "all", rows where the selected amount is NA are dropped. All amount columns are retained in the output in every case. Note that toa_amount is not reported for FY2000, so an amount filter with spend_type = "all" or "toa" excludes that year.
min_amount	Numeric. Minimum spending amount in thousands of current US dollars. Rows below this threshold are excluded. Use NA (the default) for no lower bound.
max_amount	Numeric. Maximum spending amount in thousands of current US dollars. Rows above this threshold are excluded. Use NA (the default) for no upper bound.
include_requests	Logical. Each C-1 report covers two or three fiscal years, the last of them as a budget request, and the data holds every fiscal year from the most recent report that covers it. That is an enacted or actual figure for every year except FY2016, whose only report among the source files is the one that requests it. FY2010 also has 26 lines that come from an overseas contingency operations request sheet and are on no enacted sheet. TRUE (the default) returns everything, with those rows flagged by the is_request column. FALSE drops FY2016 and the 26 FY2010 lines.
startyear	Numeric. The first fiscal year for the series.
endyear	Numeric. The last fiscal year for the series.

Details

Each fiscal year is taken whole from the most recent C-1 report that covers it, so a project-year appears once. A row is one line of that report; source_file, source_sheet and source_row identify it.

Several fields appear only in a subset of years, because the C-1 exhibit layout changed over time and a later report does not always print what an earlier one did. transaction_type and appn_title are FY2001-FY2002 only; pe FY2003-FY2004; dollar_type and existing_mission FY2003-FY2005; location_code FY2000-FY2006; facility_category_code FY2000 and FY2003-FY2006; project_number FY2000-FY2006 and FY2013 onward; state_country_sort FY2006; classification FY2011 onward; facility_category_title FY2003-FY2005, FY2007-FY2021 and FY2026. FY2006 reports appropriations only: toa_amount, auth_amount, auth_appn_amount and budget_activity_title are missing for that year, and toa_amount for FY2000. percent_recapitalization has no values. Filtering on one of these implicitly restricts the result to the years that report it.

The *_reported columns (organization_reported, budget_activity_title_reported, facility_category_title_reported, location_name_reported, project_title_reported) preserve the label exactly as printed in the source workbook, before harmonization. source_file, source_sheet, source_row, report_year, and sheet_type record provenance and can be used to trace any row back to the originating exhibit.

latitude and longitude are present only where the location was found inside the country or state the row is filed under. Rows filed under no country (unspecified, worldwide and classified locations) and rows whose location is a label rather than a place ("Various Locations") have none. geo_source says where a point came from: a geocoding service, or "manual" for the installations the services placed at the centre of the country or at another place of the same name, which are entered by hand.

Value

A data frame of location-project-year observations of U.S. military construction spending. Amount fields are in thousands of current US dollars.

Author(s)

Michael E. Flynn

References

Michael A. Allen, Michael E. Flynn, and Carla Martinez Machain. 2020. "Outside the wire: US military deployments and public opinion in host states." *American Political Science Review*. 114(2): 326-341.

Examples

```
## Not run:
library(troopdata)

# All observations from 2008 through 2019
example <- get_bulddata(startyear = 2008, endyear = 2019)

# Kuwait and Japan, budget transactions only. transaction_type is reported
# for FY2001 and FY2002 only.
example2 <- get_bulddata(
  host = c("KWT", "JPN"),
  transaction_type = "BUDGET",
  startyear = 2001,
  endyear = 2002
)

# Army and Navy major construction, TOA >= $10 million
example3 <- get_bulddata(
  organization = c("Army", "Navy"),
  budget_activity = "MAJOR CONSTRUCTION",
  spend_type = "toa",
  min_amount = 10000,
  startyear = 2015,
  endyear = 2023
)
```

```

)

# Projects matching "hospital" in title
example4 <- get_bulddata(
  project = "hospital",
  spend_type = "appn",
  startyear = 2010,
  endyear = 2023
)

# Enacted aircraft maintenance construction, excluding budget requests
example5 <- get_bulddata(
  facility_category = "Aircraft maintenance facilities",
  include_requests = FALSE,
  startyear = 2010,
  endyear = 2024
)

## End(Not run)

```

get_exercises

Function to retrieve customized multilateral military exercise data

Description

get_exercises() generates a customized data frame containing exercise-country-year observations of multilateral military exercises. Users can subset the data by participating country, year, exercise duration, geographic location, exercise name, the domain(s) of the exercise (e.g., air, land, sea), the mission focus (warfighting, humanitarian, peacekeeping), and the number of participating countries.

Usage

```

get_exercises(
  country = NULL,
  startyear = NULL,
  endyear = NULL,
  min_duration = NULL,
  max_duration = NULL,
  location = NULL,
  exercise_name = NULL,
  domain = NULL,
  focus = NULL,
  min_participants = NULL,
  max_participants = NULL
)

```

Arguments

country	The Gleditsch and Ward (G&W) numeric country code or country name for the participating country or countries to include. Numeric input is matched exactly against the gwcode column. Character input is matched against the country column using a case-insensitive grepl fuzzy match, so partial names are accepted (e.g., "korea" returns both Koreas). Multiple values can be supplied as a vector. The default is NULL, which returns all participating countries.
startyear	The first year for the series. The default is set to the minimum year in the currently published data.
endyear	The last year for the series. The default is the maximum year in the currently published data.
min_duration	Numeric. Minimum exercise duration in days (inclusive). Default is NULL (no minimum filter).
max_duration	Numeric. Maximum exercise duration in days (inclusive). Default is NULL (no maximum filter).
location	Character. A string or vector of strings used to subset exercises by geographic location. Matched against the Location column with a case-insensitive grepl fuzzy match. Default is NULL.
exercise_name	Character. A string or vector of strings used to subset exercises by name. Matched against both the Ex_Name and Series_Name columns with a case-insensitive grepl fuzzy match (e.g., "cobra" matches "Cobra Gold"). Default is NULL.
domain	Character. A string or vector of strings indicating one or more exercise domains (warfighting environments) to include. Accepted values are "air", "land", "sea", "amphibious", and "cyber". Matching is case-insensitive. An exercise is returned if it is flagged for any of the supplied domains (logical OR). Default is NULL, which returns all domains.
focus	Character. A string or vector of strings indicating one or more mission focuses to include. Accepted values are "warfighting", "humanitarian", and "peacekeeping". Matching is case-insensitive. An exercise is returned if it is flagged for any of the supplied focuses (logical OR). Default is NULL, which returns all mission focuses.
min_participants	Numeric. Minimum number of participating countries in the exercise (inclusive). Default is NULL (no minimum filter).
max_participants	Numeric. Maximum number of participating countries in the exercise (inclusive). Default is NULL (no maximum filter).

Value

get_exercises() returns a data frame containing exercise-country-year observations of multilateral military exercises that match the specified filter criteria.

Author(s)

Michael E. Flynn

References

D’Orazio, Vito; Galambos, Kevin, 2021, "Multinational Military Exercises, 1980-2010", <https://doi.org/10.7910/DVN/KHFC> Harvard Dataverse, V1.

Gleditsch, Kristian S., and Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25(4): 393-413.

Examples

```
## Not run:
library(tidyverse)
library(troopdata)

# Pull all exercises that include South Korea between 2000 and 2015.
korea_exercises <- get_exercises(country = "korea",
                                startyear = 2000,
                                endyear = 2015)

# Pull all naval and amphibious exercises lasting at least 5 days.
sea_exercises <- get_exercises(domain = c("sea", "amphibious"),
                               min_duration = 5)

# Pull all "Cobra Gold" exercises in Thailand.
cobra_gold <- get_exercises(exercise_name = "cobra gold",
                           location = "thailand")

# Pull large-scale humanitarian exercises (10 or more participants).
large_hadr <- get_exercises(focus = "humanitarian",
                           min_participants = 10)

## End(Not run)
```

get_troopdata

Function to retrieve customized U.S. troop deployment data

Description

get_troopdata() generates a customized data frame of U.S. military personnel by location and year, for every country, territory and other location in which the Department of Defense reports personnel, the United States included. It can also return the figures by report period (quarters), the figures for the fifty states and the District of Columbia (state_data), and the rows of the underlying reports (reports).

Usage

```
get_troopdata(
  host = NULL,
  branch = FALSE,
  startyear = 1950,
```

```

endyear = 2026,
quarters = FALSE,
guard_reserve = FALSE,
civilians = FALSE,
state_data = FALSE,
reports = FALSE,
afloat = c("exclude", "include", "separate")
)

```

Arguments

host A Gleditsch and Ward numeric country code, an ISO3C code, or a country name, for the host country or countries in the series. Territories and other locations that the Gleditsch and Ward list does not cover carry the package's own codes: Puerto Rico is 6, Greenland 1002 and Guam 1008, and `unique(troopdata_rebuild_long[, c("ccode", "countryname")])` lists them all. These are not Correlates of War codes: Germany is 260 here, not 255. The values given must all be of one kind; a value that matches nothing is left out, and a warning names it. A region name can be given instead; see below. With `state_data = TRUE`, give a state name or a numeric FIPS code. State names are matched without regard to case: a full name returns that state alone ("kansas" is Kansas, and "Virginia" is not also West Virginia), and part of a name returns every state whose name contains it. The default, `NULL`, returns every location, and so does `NA`. A missing value inside a vector of hosts is dropped with a warning.

A region is asked for by name. `host` is matched, as a substring and ignoring case, against the `region` column (`unique(get_troopdata())$region`) lists the names), and every location in a matching region is returned, one row per location and period, exactly as a query with no `host` returns them. Nothing is added up; the `region` column says which region each row belongs to. A string that matches several regions returns the locations of each: "Asia" matches "East Asia & Pacific", "South Asia" and "Europe & Central Asia". A string that matches a country name is read as a country, so "Africa" returns South Africa and the Central African Republic and "America" returns American Samoa; give enough of the region's name to tell them apart ("Sub-Saharan Africa", "North America"). In both of these cases a warning says what the string matched. A full region name, or a string that matches one region only, raises none.

A country name is matched as a substring, so a query reaches every historical form of a divided country: "Korea" returns North and South, "Vietnam" returns North, South and the unified state, "Czech" returns Czechoslovakia and the Czech Republic. "Germany" returns both the Federal Republic (260) and the German Democratic Republic (265); "West Germany" and "East Germany" reach them individually. Where a name spans several country codes the rows are returned separately, one per code and period, and a warning says which codes matched – aggregate across `ccode` for a single national series.

Note that matching on ISO3C codes cannot reach every location in the data. A few locations have no ISO3C code and are left out of any ISO3C filter: at present the British West Indies (1950-1965), the Spratly Islands and Kashmir. Passing `host` as a country name or a Gleditsch and Ward country code reaches them,

and a warning names them and gives the current count whenever an ISO3C filter covers years in which they have rows. Most territories and historical states do have a code: Greenland is "GRL", Bermuda "BMU", Diego Garcia "IOT" and Yugoslavia "YUG". A unit with no code of its own carries its parent's: the Azores carry Portugal's, "PRT". Locations that share a code are returned on rows of their own, so host = "PRT" returns Portugal and the Azores separately, each with its own country code. Add them up to get a total for the ISO3C code.

Personnel afloat have a location of their own, named "Afloat" (country code 10200, no ISO3C code, region "Afloat"), with values for 1950 and 1953 through 2007. It is returned with everything else when host is left empty; host = "Afloat" or host = 10200 returns it alone. It is not a country, so leave it out when adding up countries. See afloat for what it holds.

branch	Logical. Should the function return the total only, or the total and the active duty figures for the individual branches (army_ad, navy_ad, air_force_ad, marine_corps_ad, coast_guard_ad and space_force_ad)? Default is FALSE.
startyear	The first year for the series. The default is set to 1950. The state data begin in 2008.
endyear	The last year for the series. The default is the maximum year in the currently published data. The latest year holds only the reports published so far, so its annual figures can rest on fewer reports than those of a full year.
quarters	Logical. Should the function return the figures of each report rather than one row per year? If FALSE, the default, each row is a location and year, and every column holds the largest value reported for that location in that year. If TRUE, each row is a location and report period, with month and quarter columns, and its figures are those of that one report. See "How values are handled".
guard_reserve	Logical. Should the function return values for the National Guard and Reserve, and troops_all, the total of active duty, guard and reserve personnel? The reports give guard and reserve figures from 2008; they are missing (NA) for earlier years, where troops_all equals troops_ad. Default is FALSE.
civilians	Logical. Should the function return values for civilian DoD personnel? The reports give them from 2008, and they are missing (NA) for earlier years. Default is FALSE.
state_data	Logical. Should the function return the data for the fifty U.S. states and the District of Columbia rather than for countries? These data begin in 2008, and host is then a state name or a numeric FIPS code. Default is FALSE.
reports	Logical. Should the function return reports for the specified countries and years? Default is FALSE. The reports are returned as reported, without aggregation, so a country code may appear on more than one row in a period where the report breaks that country's territories out separately. Requires quarters = TRUE.
afloat	How to treat Navy and Marine Corps personnel afloat in the reports for 1950 through 2007. One of "exclude" (the default), "include" or "separate". Those reports count personnel ashore and personnel afloat separately, and attribute some of the personnel afloat to a location: to the country of the nearest port or the country whose line they are printed on (1953-1976), and, for the fleet in home waters, to the United States (1950, 1953-2007; from 1968 the reports

give that figure for the United States and its territories together). The rest are given for a region or for the world only.

- "exclude": every location holds personnel ashore. troops_ad, navy_ad and marine_corps_ad leave out anyone afloat, and the "Afloat" location holds everyone afloat worldwide.
- "include": personnel afloat who are attributed to a location are added to that location's troops_ad, navy_ad and marine_corps_ad. The "Afloat" location holds only those attributed to no location.
- "separate": the figures ashore are as for "exclude", and the personnel afloat attributed to each location are returned beside them in troops_afloat, navy_afloat and marine_corps_afloat. navy_ad and marine_corps_ad are returned as well, whatever branch is set to, so the two can be compared. The "Afloat" location holds only those attributed to no location.

Added up over every row, the three give the same total; they differ in where the personnel afloat are put. In the three afloat columns a zero means a report that lists personnel afloat by location has none for that one (1953-1976). A missing value means the report gives no figure for the location: 1950 and 1977-2007 for everywhere but the United States (and Cuba in 1950, where the report puts 171 Marines afloat), 1951 and 1952, and every year from 2008. For 1954-1956 the figure also holds mobile units temporarily based ashore, which those three reports do not separate.

From September 2008 the DMDC reports count a ship's crew at its home port, inside the state or country figure, and there is no afloat figure at all, so the argument changes nothing from 2008 on and has no effect on state_data = TRUE. Navy figures on either side of 2008 are closest to comparable with afloat = "include".

Value

A data frame with one row per location and year, or one row per location and report period when quarters = TRUE. Each row carries the location's identifiers (ccode, iso3c, countryname and region, or state and fipscode for the state data), the year (with month and quarter when quarters = TRUE), troops_ad, the number of active duty personnel, and the columns asked for with branch, guard_reserve, civilians and afloat. With reports = TRUE the columns are those of [troopdata_rebuild_reports](#), with its three afloat columns only when afloat = "separate". The section "How values are handled" says what the figures are and how they are put together.

How values are handled

The figures come from the personnel reports of the Department of Defense, published today by the Defense Manpower Data Center (DMDC). The rules below say how the reports become the values this function returns. The "Rebuild Notes" article on the package website gives more background.

Annual values are the highest value reported in the year. With quarters = FALSE every column holds the largest value reported for that location in that year. It is not an average, and it is not the figure of one fixed month. The data have one report a year through 2012 (dated June for 1950 and 1953-1956, and September from 1957), two for 2013 (September and December) and four a year from 2014 (March, June, September and December), so the rule matters mainly from 2013 on. The latest year holds only the reports published so far.

The maximum is taken for each column separately, so `troops_ad` can come from one report and `army_ad` or `navy_ad` from another. The branch columns of an annual row therefore need not add up to its `troops_ad`, and `troops_all` need not equal `troops_ad` plus the guard and reserve columns. For Japan in 2019 the four reports give 56,134, 55,327, 55,245 and 57,094 active duty personnel. The annual `troops_ad` is 57,094, the December figure, while the annual `army_ad`, 2,671, is the June figure and the annual `navy_ad`, 20,846, is the March figure. Use `quarters = TRUE` for figures that all come from one report.

Totals. `troops_ad` is active duty personnel. In the country data a row that comes from a report has a `troops_ad` equal to the sum of its branch columns. The state data keep the total each report prints, which is lower than the sum of the branch columns for some states in June 2021, December 2021 and March 2022, so the states do not add up to the United States figure in those three quarters. `troops_all`, returned with `guard_reserve = TRUE`, is `troops_ad` plus the seven National Guard and Reserve columns. The United States figure is the fifty states and the District of Columbia. Puerto Rico, Guam and the other territories are locations of their own and are not part of it.

Zeros. A country that a report does not list has 0 for that period. A zero therefore means either that the report prints zero or that it does not list the country. Territories and the other locations that carry the package's own codes are treated differently, as are a few small states in the years before their independence: they have a row only for the periods in which a report lists them, so an unlisted territory is absent from the result rather than zero.

Missing values. NA means that the reports give no figure of that kind: the guard, reserve and civilian columns and `coast_guard_ad` before 2008, `space_force_ad` before September 2023, the branch columns for 1951 and 1952, and the afloat columns as described under `afloat`. In annual output a value is missing only if it is missing in every report of that year. Otherwise it is the largest of the values reported.

Figures that are not taken from a report.

- 1951 and 1952 have no report. `troops_ad` for those years moves in equal steps from the June 1950 figure to the June 1953 figure, and the branch columns are missing. This is done for countries. Most territories and other locations with the package's own codes, the "Afloat" location among them, have no rows for those two years.
- The Army reported nothing for December 2022, March 2023 and June 2023. In the country data `army_ad`, `army_national_guard` and `army_reserve` for those quarters move in equal steps from the September 2022 figure to the September 2023 figure, and the totals are those figures plus what the other services reported. The state data and the reports (`reports = TRUE`) are left as published, with the Army figures and the totals missing for those three quarters. The one exception is the United States row of the reports, which is not a line of a report. For those quarters it carries a fixed Army figure of 404,114 and zeros, not missing values, in the Army guard and reserve columns.
- The June 2023 report prints its civilian columns, and five of its guard and reserve columns, one row low for part of the overseas list, so that Qatar's line holds Puerto Rico's figures and Uruguay's the United Kingdom's. The country data take each figure from the line below the one it is printed on. `reports = TRUE` shows the report as published.
- Where the reports do not give the personnel deployed to a war zone, a figure from Kane's data or from public reporting is used: Afghanistan for 2001-2005 and 2018-2020, Iraq for 2003-2007 and 2018-2021, Kuwait for 2003-2007 and Syria for 2018-2021. These are totals only, with zeros in the branch columns. From December 2017 the reports leave out personnel deployed to Afghanistan, Iraq and Syria, and no outside figure is used after those years, so the

values for Afghanistan from 2021 and for Iraq and Syria from 2022 are the handful of personnel the reports print and understate the U.S. presence. With `quarters = TRUE`, the outside figures for Iraq in 2018-2021 and for Syria in 2021 are on the June row only, and the other quarters of those years hold what the report prints, which is zero or close to it.

- The reports for 1957 through 2013 are dated September (and December in 2013). A row dated June in those years is not a report. It holds a figure from Kane's data, or one of the figures above, and is kept only for a location and year in which no report gives a figure or the report prints a zero in its place. Nearly all of these rows are zero.

The source column of `troopdata_rebuild_long` says where a figure comes from, with two exceptions: the rows of the three Army quarters carry the label of their report, and the Iraq figure of 5,200 for June 2018, 2019 and 2020 has no label.

Where personnel are counted. Personnel are counted at the location the report lists them under. Personnel that a report lists under no location (lines such as "Transients", "Undistributed", "Departmental Headquarters" and "Unknown") are not in the data, so the rows do not add up to the worldwide total the report prints. Where a report lists two places under one country code (Japan and the Ryukyu Islands through 1973, for example) the country figure is their sum, and `reports = TRUE` shows each row. For 1950 through 2007 the figures are personnel ashore, and personnel afloat are handled by the `afloat` argument. From September 2008 the reports count a ship's crew at its home port, inside the country or state figure. From December 2015 to December 2017 the reports list between 77,000 and 105,000 personnel, nearly all Navy, in the United States block but in no state. They cannot be traced to a state or a country and are in none of the figures, so the United States total and its Navy figure are lower in those quarters.

Author(s)

Michael E. Flynn

References

Tim Kane. Global U.S. troop deployment, 1950-2003. Technical Report. Heritage Foundation, Washington, D.C.

Michael A. Allen, Michael E. Flynn, and Carla Martinez Machain. 2022. "Global U.S. military deployment data: 1950-2020." *Conflict Management and Peace Science*. 39(3): 351-370.

Examples

```
## Not run:
library(tidyverse)
library(troopdata)

example <- get_troopdata(host = "United States",
                        branch = TRUE,
                        startyear = 1980,
                        endyear = 2015)

head(example)

## End(Not run)
```

mme_long

*Multilateral Military Exercises (MME) data, long format***Description**

mme_long returns a data frame containing exercise-country-year observations of multilateral military exercises. Built from the MME version 7 data ([doi:10.7910/DVN/KHFODX](https://doi.org/10.7910/DVN/KHFODX)) and reshaped so each row represents a single participating country in a single year of a single exercise. This is the data object underlying get_exercises().

Usage

mme_long

Format

A data frame with exercise-country-year observations including the following variables:

MMEID Unique exercise identifier from the MME source data.

Ex_Name The name of the individual exercise (e.g., "Cobra Gold 93").

Series_Name The name of the broader exercise series the exercise belongs to (e.g., "Cobra Gold").

gwcode Numeric Gleditsch and Ward country code for the participating country. Looked up from country via the countrycode package; NA for non-country participants such as "NATO" or regional groupings.

country Character vector of participating country names as recorded in the MME source data.

year The year of the observation. Exercises spanning multiple years are expanded so that each year between s.year and e.year produces its own row.

Location The geographic location where the exercise was held (free-text from the source data).

lat Latitude of the exercise location.

lon Longitude of the exercise location.

StartDate Original start-date string from the source data.

s.year Numeric year the exercise began.

s.month Numeric month the exercise began.

s.day Numeric day the exercise began (may be "xx" when unknown).

EndDate Original end-date string from the source data.

e.year Numeric year the exercise ended.

e.month Numeric month the exercise ended.

e.day Numeric day the exercise ended (may be "xx" when unknown).

CPX Binary indicator: command post exercise.

Air Binary indicator: air domain.

Land Binary indicator: land domain.

Sea Binary indicator: sea domain.

Amphibious Binary indicator: amphibious domain.

Cyber Binary indicator: cyber domain.

Warfighting Binary indicator: warfighting focus.

Peacekeeping Binary indicator: peacekeeping focus.

Humanitarian Binary indicator: humanitarian focus.

FocusDescription Free-text description of the exercise's focus from the source data.

AdditionalParticipantInfo Free-text notes about participants from the source data.

participant_count Total number of participating countries in the exercise. The same value is repeated across all rows that share an MMEID. Used by the `min_participants` and `max_participants` arguments of `get_exercises()`.

Value

Returns the full data frame of exercise-country-year observations of multilateral military exercises from 1980 forward.

Source

D'Orazio, Vito; Galambos, Kevin, 2021, "Multinational Military Exercises, 1980-2010", doi:[10.7910/DVN/KHFODX](https://doi.org/10.7910/DVN/KHFODX), Harvard Dataverse, V1.

troopdata_rebuild_long

U.S. overseas troop deployment data

Description

troopdata_rebuild_long is the data frame behind `get_troopdata`: U.S. military personnel by location and report period, for every country, territory and other location in the reports, the United States included.

Usage

troopdata_rebuild_long

Format

A data frame with one row per location and report period (a year and a quarter), including the following variables:

`ccode` A numeric vector of Gleditsch and Ward country codes. Territories and other locations that the Gleditsch and Ward list does not cover carry the package's own codes (Puerto Rico 6, Greenland 1002, Guam 1008 and others, all above 1000 except Puerto Rico). These are not Correlates of War codes: Germany is 260, not 255.

iso3c A character vector of ISO three character country codes. A location with no code of its own carries its parent's (the Azores carry "PRT"), and a few locations have none.

countryname A character vector of country names.

region Geographic region, using the names of the World Bank regions. Countries take the region the countrycode package gives them, territories and other locations with the package's own codes are assigned one by the package, and "Afloat" and "Antarctica" are regions of their own.

year The year of the observation.

month The month of the report: March, June, September or December.

quarter The quarter of the observation, 1 through 4.

year_quarter The year and quarter of the observation.

source Where the row's figures come from: the month and year of the DMDC report, "Kane 2006" for a figure from Kane's data, "Stepwise Imputation" for 1951 and 1952, or a label beginning "Estimate" for a figure from public reporting. Missing where no report lists the location for that period, which is where the zeros described in Details come from; 17 rows with no source are not zero, the Iraq figure of 5,200 for June 2018, 2019 and 2020 among them. The rows of December 2022, March 2023 and June 2023 carry the label of their report although their Army figures are filled in.

troops_ad The total number of active duty US military personnel deployed to the host country. In a row that comes from a report it equals the sum of the branch columns.

troops_all The total number of US military personnel deployed to the host country including guard and reserve: troops_ad plus the seven National Guard and Reserve columns. Equal to troops_ad before 2008, when the reports give no guard or reserve figures.

army_ad Total number of active duty Army personnel deployed to the host country.

navy_ad Total number of active duty Navy personnel deployed to the host country.

air_force_ad Total number of active duty Air Force personnel deployed to the host country.

space_force_ad Total number of active duty Space Force personnel deployed to the host country.

marine_corps_ad Total number of Marine Corps personnel deployed to the host country.

coast_guard_ad Total number of Coast Guard personnel deployed to the host country.

troops_afloat Navy and Marine Corps personnel afloat that a report attributes to the location, 1950 through 2007. Not included in troops_ad. See Details.

navy_afloat Navy personnel afloat that a report attributes to the location. Not included in navy_ad.

marine_corps_afloat Marine Corps personnel afloat that a report attributes to the location. Not included in marine_corps_ad.

total_selected_reserve Total number of reserve US military personnel deployed to the host country. This and the other guard, reserve and civilian columns are reported from 2008 and are missing for earlier years.

army_reserve Total number of reserve Army personnel deployed to the host country.

navy_reserve Total number of reserve Navy personnel deployed to the host country.

air_force_reserve Total number of reserve Air Force personnel deployed to the host country.

marine_corps_reserve Total number of reserve Marine Corps personnel deployed to the host country.

`coast_guard_reserve` Total number of reserve Coast Guard personnel deployed to the host country.

`army_national_guard` Total number of Army National Guard personnel deployed to the host country.

`air_national_guard` Total number of Air National Guard personnel deployed to the host country.

`army_civilian` Total number of Army civilian personnel deployed to the host country.

`navy_civilian` Total number of Navy civilian personnel deployed to the host country.

`air_force_civilian` Total number of Air Force civilian personnel deployed to the host country.

`marine_corps_civilian` Total number of Marine Corps civilian personnel deployed to the host country.

`dod_civilian` Total number of Department of Defense civilian personnel deployed to the host country.

`total_civilian` Total number of civilian personnel deployed to the host country.

Details

Report periods. Each row holds the figures of one period. The data have one report a year through 2012 (dated June for 1950 and 1953-1956, and September from 1957), two for 2013 (September and December) and four a year from 2014 (March, June, September and December), and they stop at the latest report. [get_troopdata](#) returns these rows with `quarters = TRUE`. Its annual figures (`quarters = FALSE`) are the largest value of each column over the periods of a year, taken column by column, so they are not stored here and the columns of an annual row need not come from the same report.

Zeros and missing values. A country that a report does not list has 0 for that period and no source. A zero therefore means either that the report prints zero or that it does not list the country. Territories and the other locations that carry the package's own codes, and a few small states in the years before their independence, have a row only for the periods in which a report lists them, so the data are not a balanced panel for them. NA means that the reports give no figure of that kind: the guard, reserve and civilian columns and `coast_guard_ad` before 2008, `space_force_ad` before September 2023, the branch columns for 1951 and 1952, and the afloat columns as described below.

Figures that are not taken from a report. The source column marks them, apart from the two cases noted under source above.

- 1951 and 1952 have no report. `troops_ad` for those years moves in equal steps from the June 1950 figure to the June 1953 figure, and the branch columns are missing. This is done for countries. Most territories and other locations with the package's own codes, the "Afloat" location among them, have no rows for those two years.
- The Army reported nothing for December 2022, March 2023 and June 2023. `army_ad`, `army_national_guard` and `army_reserve` for those quarters move in equal steps from the September 2022 figure to the September 2023 figure, and `troops_ad`, `total_selected_reserve` and `troops_all` are those figures plus what the other services reported. The other columns are as reported. [troopdata_rebuild_reports](#) and [troopdata_rebuild_us_states](#) keep those quarters as published, with the Army figures and the totals missing.
- The June 2023 report prints part of its overseas list one row low: the civilian columns from Montenegro to the end of the list, and the Navy Reserve, Marine Corps Reserve, Air National Guard, Air Force Reserve and Coast Guard Reserve columns from Morocco to Wake

Island. Qatar's line holds Puerto Rico's 2,201 civilians and Uruguay's the United Kingdom's 1,383. The figures here are read from the line below the one they are printed on, which is where the March and September 2023 reports and the report's own totals put them. [troopdata_rebuild_reports](#) keeps the report as published.

- Where the reports do not give the personnel deployed to a war zone, a figure from Kane's data or from public reporting is used: Afghanistan for 2001-2005 and 2018-2020, Iraq for 2003-2007 and 2018-2021, Kuwait for 2003-2007 and Syria for 2018-2021. These are totals only, with zeros in the branch columns. From December 2017 the reports leave out personnel deployed to Afghanistan, Iraq and Syria, and no outside figure is used after those years, so the values for Afghanistan from 2021 and for Iraq and Syria from 2022 are the handful of personnel the reports print and understate the U.S. presence. The outside figures for Iraq in 2018-2021 and for Syria in 2021 are on the June row only, and the other quarters of those years hold what the report prints, which is zero or close to it.
- The reports for 1957 through 2013 are dated September (and December in 2013). A row dated June in those years is not a report. It holds a figure from Kane's data, or one of the estimates above, and is kept only for a location and year in which no report gives a figure or the report prints a zero in its place. Nearly all of these rows are zero.

Locations. Personnel are counted at the location the report lists them under. Personnel that a report lists under no location (lines such as "Transients", "Undistributed", "Departmental Headquarters" and "Unknown") are not in the data. Where a report lists two places under one country code (Japan and the Ryukyu Islands through 1973, for example) the row is their sum. The United States is the fifty states and the District of Columbia. Puerto Rico, Guam and the other territories are locations of their own.

Personnel ashore and afloat. Before 2008 `troops_ad`, `navy_ad` and `marine_corps_ad` hold personnel ashore for every location, the United States included. The reports for 1950 through 2007 count Navy and Marine Corps personnel afloat separately, and attribute some of them to a location: to the country of the nearest port or the country whose line they are printed on (1953-1976), and, for the fleet in home waters, to the United States (1950, 1953-2007; from 1968 the reports give that figure for the United States and its territories together). Those are in `troops_afloat`, `navy_afloat` and `marine_corps_afloat`, beside the ashore figures and not inside them. A zero there means a report that lists personnel afloat by location has none for that one (1953-1976); a missing value means the report gives no figure for the location (everywhere but the United States in 1950 and 1977-2007, apart from 171 Marines afloat at Cuba in 1950; the years 1951 and 1952; and everything from 2008). For 1954-1956 the figure also holds mobile units temporarily based ashore, which those three reports do not separate.

Everyone afloat is also carried once, worldwide, in a location named "Afloat" (ccode 10200, no ISO code, region "Afloat"), with values for 1950 and 1953 through 2007. It is not a country, and it includes the personnel the three afloat columns attribute to locations: `troops_ad` added up over every row, that one included, counts everyone the data place somewhere once, and adding the afloat columns to it as well would count those personnel twice. (It is still below the worldwide total a report prints, which also includes the personnel it lists under no location.) [get_troopdata](#) handles this through its `afloat` argument, which leaves the attributed personnel out of the locations, adds them in, or returns them separately, and adjusts the "Afloat" location to match. From September 2008 the DMDC reports count a ship's crew at its home port, inside the state or country figure, and there is no afloat figure, so the ashore Navy figures for countries and states that are home ports step up between 2007 and 2008.

The reports did not keep to that from December 2015 to December 2017. They list 77,120 personnel under "Unknown" in the United States block in December 2015, and between 88,500 and 104,703 in each report from March 2016 to December 2017 under "Armed Forces Europe", "Armed Forces Pacific" and "Armed Forces the Americas", nearly all of them Navy. Those personnel cannot be traced to a state or a country and are in none of the figures here, the "Afloat" location included. The United States total and its Navy figure are therefore lower in those quarters and return in March 2018: the annual Navy figure for the United States is 278,123 in 2015, 188,184 in 2016, 198,679 in 2017 and 289,853 in 2018. Japan's Navy figure for December 2015 (41,546, against 22,030 in September 2015 and 12,234 in March 2016) is as that report prints it, and is why the annual figure for Japan in 2015, 71,567, stands well above the years around it.

Value

Returns the full data frame containing observations of US military deployments by location (countries and territories) from 1950 through the latest report in this version of the package, March 2026.

Source

<https://www.heritage.org/defense/report/global-us-troop-deployment-1950-2005>
doi:10.1177/07388942211030885

troopdata_rebuild_reports

DMDC Deployment Reports

Description

troopdata_rebuild_reports returns a data frame containing DMDC reports on US military deployments.

Usage

```
troopdata_rebuild_reports
```

Format

A data frame with one row per line of a report, including the following variables. A country code can appear on more than one row in a period, because the rows keep the locations and names the reports print (Japan and the Ryukyu Islands are both under 740, for example).

ccode A numeric vector of Gleditsch and Ward country codes, with the package's own codes for territories and other locations the list does not cover. See [troopdata_rebuild_long](#).

iso3c A character vector of ISO three character country codes.

countryname A character vector of country names. They are tidied but not fully standardized, so a code can carry more than one name here and a name can differ from the one in [troopdata_rebuild_long](#) (260 is "Germany" here). Location holds the name as the report prints it.

region Geographic region, as in [troopdata_rebuild_long](#), but missing for a number of territories and historical locations.

year The year of the observation.

month The month of the observation.

quarter The quarter of the observation.

source The DMDC report source of the observation.

Location The geographic location listed in the DMDC reports.

Total "Total number of US military personnel deployed to the host country.

Total Ashore "Total number of US military personnel deployed to the host country, excluding those at sea.

Total Afloat "Total number of US military personnel deployed to the host country, at sea.

Army Total Total number of Army personnel deployed to the host country.

Navy Ashore Total number of Navy personnel deployed to the host country, excluding those at sea.

Navy Temporary Ashore Total number of Navy personnel deployed to the host country, temporarily.

Navy Other Total number of Navy personnel deployed to the host country, in other capacities.

Marine Corps Ashore Total number of Marine Corps personnel deployed to the host country, excluding those at sea.

Marine Corps Afloat Total number of Marine Corps personnel deployed to the host country, at sea.

Air Force Total Total number of Air Force personnel deployed to the host country.

Navy Afloat Total number of Navy personnel deployed to the host country, at sea.

Navy Total Total number of Navy personnel deployed to the host country.

Marine Corps Total Total number of Marine Corps personnel deployed to the host country.

troops_ad The total number of active duty US military personnel deployed to the host country.

army_ad Total number of active duty Army personnel deployed to the host country.

navy_ad Total number of active duty Navy personnel deployed to the host country.

marine_corps_ad Total number of active duty Marine Corps personnel deployed to the host country.

space_force_ad Total number of active duty Space Force personnel deployed to the host country.

air_force_ad Total number of active duty Air Force personnel deployed to the host country.

coast_guard_ad Total number of Coast Guard personnel deployed to the host country.

troops_afloat Navy and Marine Corps personnel afloat that a report attributes to the location, 1950 through 2007. Not included in troops_ad. See Details.

navy_afloat Navy personnel afloat that a report attributes to the location. Not included in navy_ad.

marine_corps_afloat Marine Corps personnel afloat that a report attributes to the location. Not included in marine_corps_ad.

Macro Location The geographic location listed in the DMDC reports.

Army Active Duty Total number of active duty Army personnel deployed to the host country.

Navy Active Duty Total number of active duty Navy personnel deployed to the host country.

Marine Corps Active Duty Total number of active duty Marine Corps personnel deployed to the host country.

Air Force Active Duty Total number of active duty Air Force personnel deployed to the host country.

Coast Guard Active Duty Total number of active duty Coast Guard personnel deployed to the host country.

Space Force Active Duty Total number of active duty Space Force personnel deployed to the host country.

Total Active Duty Total number of active duty US military personnel deployed to the host country.

Army National Guard Total number of Army National Guard personnel deployed to the host country.

Army Reserve Total number of reserve Army personnel deployed to the host country.

Navy Reserve Total number of reserve Navy personnel deployed to the host country.

Marine Corps Reserve Total number of reserve Marine Corps personnel deployed to the host country.

Air National Guard Total number of Air National Guard personnel deployed to the host country.

Air Force Reserve Total number of reserve Air Force personnel deployed to the host country.

Coast Guard Reserve Total number of reserve Coast Guard personnel deployed to the host country.

Total Selected Reserve Total number of reserve US military personnel deployed to the host country.

Army Civilian Total number of Army civilian personnel deployed to the host country.

Navy Civilian Total number of Navy civilian personnel deployed to the host country.

Marine Corps Civilian Total number of Marine Corps civilian personnel deployed to the host country.

Air Force Civilian Total number of Air Force civilian personnel deployed to the host country.

DOD Civilian Total number of Department of Defense civilian personnel deployed to the host country.

Total Civilian Total number of civilian personnel deployed to the host country.

Grand Total Total number of US military and civilian personnel deployed to the host country.

Details

Before 2008 `troops_ad`, `navy_ad` and `marine_corps_ad` hold personnel ashore for every location, the United States included. The reports for 1950 through 2007 count Navy and Marine Corps personnel afloat separately, and attribute some of them to a location: to the country of the nearest port or the country whose line they are printed on (1953-1976), and, for the fleet in home waters, to the United States (1950, 1953-2007; from 1968 the reports give that figure for the United States and its territories together). Those are in `troops_afloat`, `navy_afloat` and `marine_corps_afloat`, beside the ashore figures and not inside them. A zero there means a report that lists personnel afloat

by location has none for that one (1953-1976); a missing value means the report gives no figure for the location (everywhere but the United States in 1950 and 1977-2007, apart from 171 Marines afloat at Cuba in 1950; the years 1951 and 1952; and everything from 2008). For 1954-1956 the figure also holds mobile units temporarily based ashore, which those three reports do not separate.

Everyone afloat is also carried once, worldwide, in a location named "Afloat" (ccode 10200, no ISO code, region "Afloat"), with values for 1950 and 1953 through 2007. It is not a country, and it includes the personnel the three afloat columns attribute to locations: `troops_ad` added up over every row, that one included, counts everyone the data place somewhere once, and adding the afloat columns to it as well would count those personnel twice. (It is still below the worldwide total a report prints, which also includes the personnel it lists under no location.) `get_troopdata` handles this through its `afloat` argument, which leaves the attributed personnel out of the locations, adds them in, or returns them separately, and adjusts the "Afloat" location to match. From September 2008 the DMDC reports count a ship's crew at its home port, inside the state or country figure, and there is no afloat figure, so the ashore Navy figures for countries and states that are home ports step up between 2007 and 2008.

The reports did not keep to that from December 2015 to December 2017. They list 77,120 personnel under "Unknown" in the United States block in December 2015, and between 88,500 and 104,703 in each report from March 2016 to December 2017 under "Armed Forces Europe", "Armed Forces Pacific" and "Armed Forces the Americas", nearly all of them Navy. Those personnel cannot be traced to a state or a country and are in none of the rows here, so the United States figures are lower in those quarters and return in March 2018.

The derived columns, whose names are in lower case, follow that rule. The columns that carry a report's own figures, such as `Total`, `Total Afloat`, `Navy Afloat` and `Navy Other`, are left as the report prints them; `navy_afloat` is built from `Navy Afloat` or `Navy Other`, whichever the report of that year uses (the United States row, which is assembled from several lines, aside). For December 2022, March 2023 and June 2023 the Army reported nothing: the Army columns and `troops_ad` are missing on those rows and the other services are as reported. The United States row is the exception. It is not a line of a report (from 2008 it is the sum of the state lines), and for those three quarters it carries a fixed Army figure of 404,114, a total built on it, and zeros rather than missing values in the Army guard and reserve columns and the totals that include them. `troopdata_rebuild_long` fills the Army figures for those quarters in equal steps between September 2022 and September 2023 instead, so the two differ for the United States there.

In 46 rows of December 2021, March 2022 and June 2022 the total a report prints, which is what `troops_ad` holds here, is lower than the sum of that row's service columns. `troopdata_rebuild_long` uses the sum.

The June 2023 report prints part of its overseas list one row low: the six civilian columns from Montenegro to the end of the list, and the Navy Reserve, Marine Corps Reserve, Air National Guard, Air Force Reserve and Coast Guard Reserve columns from Morocco to Wake Island. Each of those lines holds the figures of the location above it, so Qatar shows Puerto Rico's 2,201 civilians and Puerto Rico shows 7. The rows here are as published. `troopdata_rebuild_long` reads those figures from the right line.

Value

Returns a data frame containing DMDC reports of US military deployments by location from 1950 through the latest report in this version of the package, March 2026.

Source

<https://www.heritage.org/defense/report/global-us-troop-deployment-1950-2005>
doi:10.1177/07388942211030885

troopdata_rebuild_us_states

U.S. domestic troop deployment data, by state

Description

troopdata_rebuild_us_states returns a data frame containing information on U.S. military personnel stationed in each of the 50 U.S. states and the District of Columbia. Territories such as Puerto Rico and Guam are overseas locations and are in the country data, not here. Returned by get_troopdata() when the state_data argument is set to TRUE.

Usage

```
troopdata_rebuild_us_states
```

Format

A data frame with one row per state and report period (one report a year through 2012, two in 2013 and four a year from 2014), including the following variables:

fipscode A numeric vector of U.S. Federal Information Processing Standards (FIPS) state codes.

Used as the numeric identifier when subsetting via get_troopdata(host = <numeric>, state_data = TRUE).

state A character vector of U.S. state names. Matched without regard to case when subsetting via get_troopdata(host = <character>, state_data = TRUE): a full name returns that state alone ("kansas" is Kansas, "Virginia" is not also West Virginia), and part of a name returns every state whose name contains it ("Carolina").

year The year of the observation.

month The month of the observation.

quarter The quarter of the observation.

troops_ad The total number of active duty US military personnel stationed in the state, as the report prints it. For 95 state-quarters in June 2021, December 2021 and March 2022 the printed total is lower than the sum of the branch columns. The country data use the sum for the United States in those quarters, so the states add up to 23, 4,330 and 6,046 fewer personnel than the United States figure there. Missing in December 2022, March 2023 and June 2023, when the Army did not report.

army_ad Total number of active duty Army personnel stationed in the state.

navy_ad Total number of active duty Navy personnel stationed in the state.

air_force_ad Total number of active duty Air Force personnel stationed in the state.

marine_corps_ad Total number of active duty Marine Corps personnel stationed in the state.

coast_guard_ad Total number of active duty Coast Guard personnel stationed in the state.
 space_force_ad Total number of active duty Space Force personnel stationed in the state.
 army_national_guard Total number of Army National Guard personnel stationed in the state.
 air_national_guard Total number of Air National Guard personnel stationed in the state.
 army_reserve Total number of Army Reserve personnel stationed in the state.
 navy_reserve Total number of Navy Reserve personnel stationed in the state.
 marine_corps_reserve Total number of Marine Corps Reserve personnel stationed in the state.
 air_force_reserve Total number of Air Force Reserve personnel stationed in the state.
 coast_guard_reserve Total number of Coast Guard Reserve personnel stationed in the state.
 total_selected_reserve Total number of reserve US military personnel stationed in the state.
 troops_all The total number of US military personnel stationed in the state including guard and reserve: troops_ad plus the seven guard and reserve components, the same definition as in the country data. Missing in December 2022, March 2023 and June 2023, when the Army did not report and troops_ad is missing here.
 army_civilian Total number of Army civilian personnel stationed in the state.
 navy_civilian Total number of Navy civilian personnel stationed in the state.
 air_force_civilian Total number of Air Force civilian personnel stationed in the state.
 marine_corps_civilian Total number of Marine Corps civilian personnel stationed in the state.
 dod_civilian Total number of Department of Defense civilian personnel stationed in the state.
 total_civilian Total number of civilian personnel stationed in the state.

Details

From December 2015 to December 2017 the DMDC reports list a large number of personnel, nearly all Navy, inside the United States block but in no state: 77,120 under "Unknown" in December 2015, and between 88,500 and 104,703 in each report from March 2016 to December 2017 under "Armed Forces Europe", "Armed Forces Pacific" and "Armed Forces the Americas". They cannot be traced to a state and are not in these figures, so the Navy and total figures of the home-port states are lower in those quarters and return in March 2018.

The Army reported nothing for December 2022, March 2023 and June 2023. In those three quarters army_ad, army_national_guard, army_reserve and the totals that include them (troops_ad, total_selected_reserve and troops_all) are missing here, and the other columns are as reported. Unlike the country data, nothing is filled in. The annual figures that [get_troopdata](#) returns for a state are the largest value of each column over the quarters of the year that have one, so for 2022 and 2023 the totals and the Army figures come from the quarters the Army reported.

Value

Returns the full data frame containing state-year-quarter observations of U.S. military personnel stationed domestically from September 2008, the first report that lists the states, through the most recent reporting period.

Source

<https://www.heritage.org/defense/report/global-us-troop-deployment-1950-2005>
[doi:10.1177/07388942211030885](https://doi.org/10.1177/07388942211030885)

Index

* **datasets**

- basedata, [3](#)
- build_data_20260918, [4](#)
- bulddata, [4](#)
- mme_long, [21](#)
- troopdata_rebuild_long, [22](#)
- troopdata_rebuild_reports, [26](#)
- troopdata_rebuild_us_states, [30](#)

* **package**

- troopdata-package, [2](#)

basedata, [3](#)

build_data_20260918, [4](#)

bulddata, [4](#)

get_basedata, [7](#)

get_bulddata, [9](#)

get_exercises, [13](#)

get_troopdata, [15](#), [22](#), [24](#), [25](#), [29](#), [31](#)

mme_long, [21](#)

troopdata (troopdata-package), [2](#)

troopdata-package, [2](#)

troopdata_rebuild_long, [20](#), [22](#), [26](#), [27](#), [29](#)

troopdata_rebuild_reports, [18](#), [24](#), [25](#), [26](#)

troopdata_rebuild_us_states, [24](#), [30](#)