Identifying Data

Data Analysis

Different scenarios you may face looking for data:

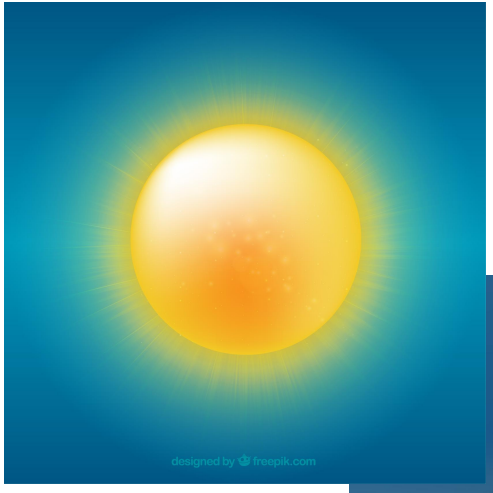
- You can find data free on the web
- Your client or supervisor provides the data to you
- You need to buy the data
- You may need to contact owners of the data to get access to them

Different types of data by how they are collected:

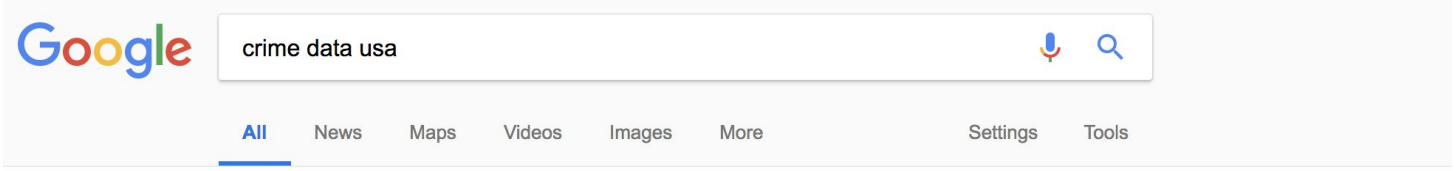
- Observational data
  - Cross-sectional data
  - Longitudinal data
  - Panel data
- Experimental data

Make sure:

- if variables are collected from multiple sources, they are merged properly
- each variable forms a column
- each observation forms a row
- each table/file stores data about one kind of observation
- column names are easy to use and informative
- obvious mistakes in the data have been removed
- missing values are formatted uniformly and correctly
- variable values are internally consistent
- appropriate transformed variables have been added



Source: freepik.com


Google
crime data usa
All
News
Maps
Videos
Images
More
Settings
Tools
About 187,000,000 results (0.46 seconds)
Crime Data Explorer
https://crime-data-explorer.fr.cloud.gov/
US flag An official website of the United States government. This site is under ... This is an open data project to improve the nation's crime data and promote ...
Explorer · Downloads & Documentation · About
Uniform Crime Reporting Statistics
https://www.ucrdatatool.gov/
Jan 26, 2017 - These data have been published each year, and since 1958, have been available in the publication Crime in the United States (CIUS).
FBI — Crime in the U.S.
https://ucr.fbi.gov/crime-in-the-u.s
Crime in the U.S.. 2017 · 2016 · 2015 · 2014 · 2013 · 2012 · 2011 · 2010 · 2009 · 2008 · 2007 · 2006 · 2005 · 2004 · 2003 · 2002 · 2001 · 2000 · 1999 · 1998 · 1997 ...
People also search for
fbi crime statistics by race
fbi ucr 2015
fbi homicide statistics by weapon
ucr 2017
crime rates over the last 20 years
fbi's crime in the united states 2016 report

U.S. DEPARTMENT OF JUSTICE • FEDERAL BUREAU OF INVESTIGATION • CRIMINAL JUSTICE INFORMATION SERVICES DIVISION

2017 JANUARY-JUNE
PRELIMINARY SEMIANNUAL UNIFORM CRIME REPORT
CRIME in the UNITED STATES

Preliminary Semiannual Uniform Crime Report, January–June, 2017

Preliminary figures indicate that law enforcement agencies throughout the nation showed an overall decrease of 0.8 percent in the number of violent crimes brought to their attention for the first 6 months of 2017 when compared with figures reported for the same time in 2016. The violent crime category includes murder, rape (revised definition), robbery, and aggravated assault. The number of property crimes in the United States from January to June of 2017 dropped 2.9 percent when compared with data for the same time period in 2016. Property crimes include burglary, larceny-theft, and motor vehicle theft. Arson is also a property crime, but data for arson are not included in property crime totals due to fluctuations in reporting. Figures for 2017 indicate that arson decreased 3.5 percent when compared with 2016 figures for the same time period.

The data presented in Tables 1 and 2 indicate the percent change in offenses known to law enforcement for the first 6 months of 2017 compared with those for the first half of 2016 by population group and region, respectively. Table 3 reflects the percent change in offenses reported within the nation for consecutive years (each year compared to the prior year). Table 4 presents the number of offenses known to law enforcement for agencies with resident populations of 100,000 or more that provided 6 months of complete data for 2017. In addition, Table 4 presents 6 months of 2016 data, where available, as a point of comparison. All data in this *Report* are preliminary.

PLEASE NOTE

In 2013, the FBI's UCR Program initiated the collection of rape data under a revised definition within the Summary Based Reporting System. The term "forcible" was removed from the offense name, and the definition was changed to "penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim."

In 2016, the FBI Director approved the recommendation to discontinue the reporting of rape data using the UCR legacy definition beginning in 2017. Therefore, the rape data reported by those agencies using the UCR legacy definition are not included in this *Report*. More information about this subject is presented in footnotes and data declarations for each table.

Caution against ranking

Figures used in this *Report* were submitted voluntarily by law enforcement agencies throughout the country. Individuals using these tabulations are cautioned against drawing conclusions by making direct comparisons between cities. Comparisons

Resources

- **Table 1**
  Percent Change, by Population Group
- **Table 2**
  Percent Change, by Region
- **Table 3**
  Percent Change, for Consecutive Years 2013-2017
- **Table 4**
  Offenses Reported to Law Enforcement, by State by City 100,000 and over in population
- **Download Table Excel Files**
  A zip folder containing the excel files for the tables listed above

Table 4

January to June 2016–2017
Offenses Reported to Law Enforcement
by State by City 100,000 and over in population

Data Declaration | Download Excel

- Alabama through California
- Colorado through Hawaii
- Illinois through Missouri
- Montana through Ohio
- Oklahoma through Wisconsin

Data Declaration

Provides the methodology used in constructing this table and other pertinent information about this table.

Table 4

January to June 2016–2017
Offenses Reported to Law Enforcement
by State by City 100,000 and over in population

Data Declaration | Download Excel

- Alabama through California
- Colorado through Hawaii
- Illinois through Missouri
- Montana through Ohio
- Oklahoma through Wisconsin

Data Declaration

Provides the methodology used in constructing this table and other pertinent information about this table.

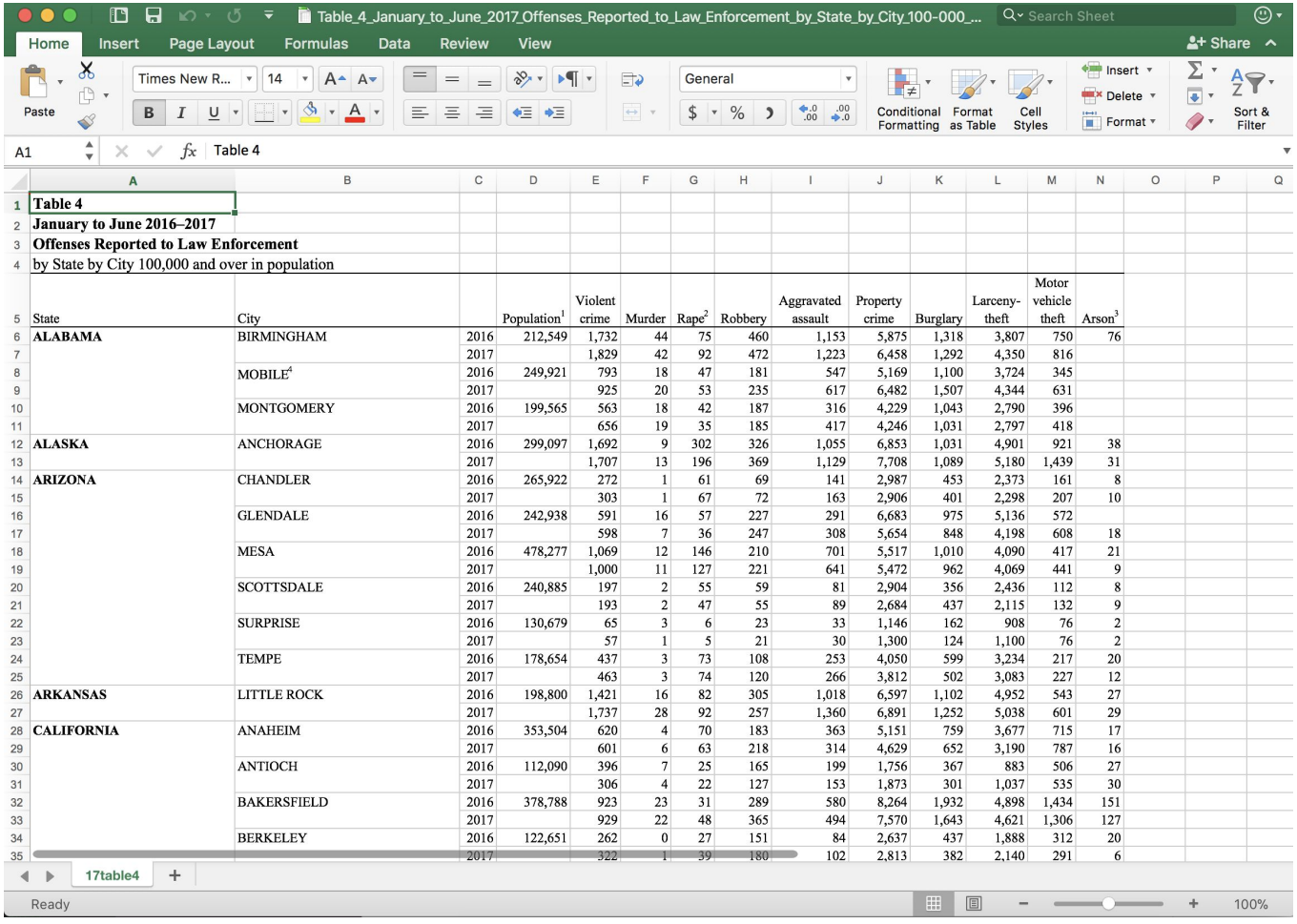


Table 4
January to June 2016–2017
Offenses Reported to Law Enforcement
by State by City 100,000 and over in population

State	City		Population[1]	Violent crime	Murder	Rape[2]	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft	Arson[3]
ALABAMA	BIRMINGHAM	2016	212,549	1,732	44	75	460	1,153	5,875	1,318	3,807	750	76
		2017		1,829	42	92	472	1,223	6,458	1,292	4,350	816	
	MOBILE[4]	2016	249,921	793	18	47	181	547	5,169	1,100	3,724	345	
		2017		925	20	53	235	617	6,482	1,507	4,344	631	
	MONTGOMERY	2016	199,565	563	18	42	187	316	4,229	1,043	2,790	396	
		2017		656	19	35	185	417	4,246	1,031	2,797	418	
ALASKA	ANCHORAGE	2016	299,097	1,692	9	302	326	1,055	6,853	1,031	4,901	921	38
		2017		1,707	13	196	369	1,129	7,708	1,089	5,180	1,439	31
ARIZONA	CHANDLER	2016	265,922	272	1	61	69	141	2,987	453	2,373	161	8
		2017		303	1	67	72	163	2,906	401	2,298	207	10
	GLENDALE	2016	242,938	591	16	57	227	291	6,683	975	5,136	572	
		2017		598	7	36	247	308	5,654	848	4,198	608	18
	MESA	2016	478,277	1,069	12	146	210	701	5,517	1,010	4,090	417	21
		2017		1,000	11	127	221	641	5,472	962	4,069	441	9
	SCOTTSDALE	2016	240,885	197	2	55	59	81	2,904	356	2,436	112	8
		2017		193	2	47	55	89	2,684	437	2,115	132	9
	SURPRISE	2016	130,679	65	3	6	23	33	1,146	162	908	76	2
		2017		57	1	5	21	30	1,300	124	1,100	76	2
	TEMPE	2016	178,654	437	3	73	108	253	4,050	599	3,234	217	20
		2017		463	3	74	120	266	3,812	502	3,083	227	12
ARKANSAS	LITTLE ROCK	2016	198,800	1,421	16	82	305	1,018	6,597	1,102	4,952	543	27
		2017		1,737	28	92	257	1,360	6,891	1,252	5,038	601	29
CALIFORNIA	ANAHEIM	2016	353,504	620	4	70	183	363	5,151	759	3,677	715	17
		2017		601	6	63	218	314	4,629	652	3,190	787	16
	ANTIOCH	2016	112,090	396	7	25	165	199	1,756	367	883	506	27
		2017		306	4	22	127	153	1,873	301	1,037	535	30
	BAKERSFIELD	2016	378,788	923	23	31	289	580	8,264	1,932	4,898	1,434	151
		2017		929	22	48	365	494	7,570	1,643	4,621	1,306	127
	BERKELEY	2016	122,651	262	0	27	151	84	2,637	437	1,888	312	20
		2017		322	1	39	180	102	2,813	382	2,140	291	6

```r
## importing the Excel data
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
crime <- read_excel("crime_data.xls", skip = 4) %>%
  rename(Year = X__1,
         Population = Population1,
         Rape = Rape2,
         Arson = Arson3) %>%
  select(-X__2, -X__3, -X__4, -X__5)

colnames(crime) <- make.names(colnames(crime), unique=TRUE)
```

```
head(crime)
```

```
## # A tibble: 6 x 14
##   State City   Year Population Violent.crime Murder  Rape Robbery
##   <chr> <chr> <dbl>      <dbl>         <dbl>  <dbl> <dbl>   <dbl>
## 1 ALAB… BIRM…  2016     212549          1732     44    75     460
## 2 <NA>  <NA>   2017         NA          1829     42    92     472
## 3 <NA>  MOBI…  2016     249921           793     18    47     181
## 4 <NA>  <NA>   2017         NA           925     20    53     235
## 5 <NA>  MONT…  2016     199565           563     18    42     187
## 6 <NA>  <NA>   2017         NA           656     19    35     185
## # ... with 6 more variables: Aggravated.assault <dbl>,
## #   Property.crime <dbl>, Burglary <dbl>, Larceny..theft <dbl>,
## #   Motor.vehicle.theft <dbl>, Arson <dbl>
```

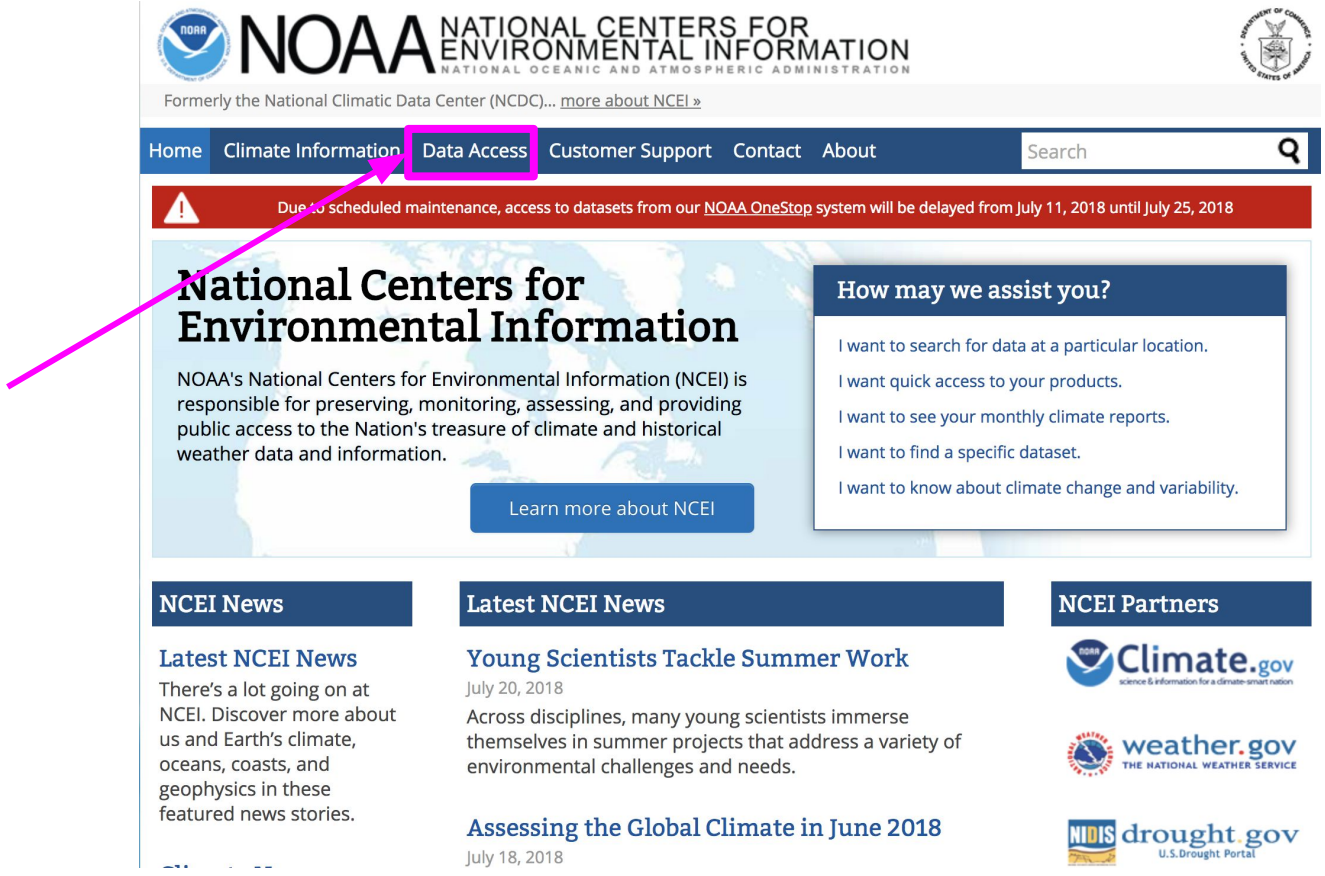
```
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

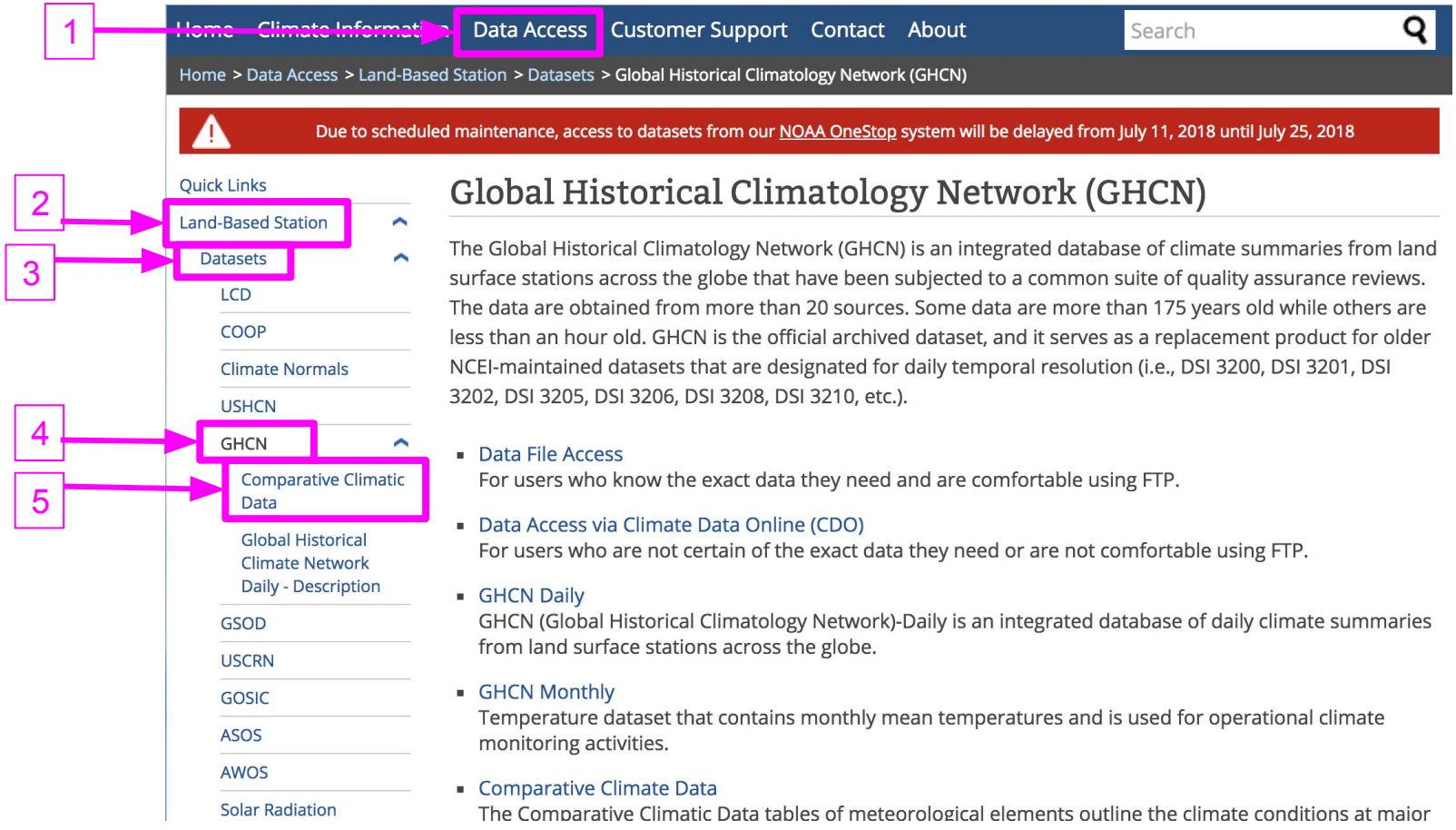
```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
crime$State <- na.locf(crime$State)
crime$City <- na.locf(crime$City)
crime$Population <- na.locf(crime$Population)
head(crime)
```

```
## # A tibble: 6 x 14
##   State City   Year Population Violent.crime Murder  Rape Robbery
##   <chr> <chr> <dbl>      <dbl>         <dbl>  <dbl> <dbl>   <dbl>
## 1 ALAB… BIRM…  2016     212549          1732     44    75     460
## 2 ALAB… BIRM…  2017     212549          1829     42    92     472
## 3 ALAB… MOBI…  2016     249921           793     18    47     181
## 4 ALAB… MOBI…  2017     249921           925     20    53     235
## 5 ALAB… MONT…  2016     199565           563     18    42     187
## 6 ALAB… MONT…  2017     199565           656     19    35     185
## # ... with 6 more variables: Aggravated.assault <dbl>,
## #   Property.crime <dbl>, Burglary <dbl>, Larceny..theft <dbl>,
## #   Motor.vehicle.theft <dbl>, Arson <dbl>
```



https://www.ncdc.noaa.gov/


1
Home
Climate Information
Data Access
Customer Support
Contact
About
Search
Home > Data Access > Land-Based Station > Datasets > Global Historical Climatology Network (GHCN)
Due to scheduled maintenance, access to datasets from our NOAA OneStop system will be delayed from July 11, 2018 until July 25, 2018
2
Quick Links
Land-Based Station
3
Datasets
LCD
COOP
Climate Normals
USHCN
4
GHCN
5
Comparative Climatic Data
Global Historical Climate Network Daily - Description
GSOD
USCRN
GOSIC
ASOS
AWOS
Solar Radiation
Global Historical Climatology Network (GHCN)
The Global Historical Climatology Network (GHCN) is an integrated database of climate summaries from land surface stations across the globe that have been subjected to a common suite of quality assurance reviews. The data are obtained from more than 20 sources. Some data are more than 175 years old while others are less than an hour old. GHCN is the official archived dataset, and it serves as a replacement product for older NCEI-maintained datasets that are designated for daily temporal resolution (i.e., DSI 3200, DSI 3201, DSI 3202, DSI 3205, DSI 3206, DSI 3208, DSI 3210, etc.).
Data File Access
For users who know the exact data they need and are comfortable using FTP.
Data Access via Climate Data Online (CDO)
For users who are not certain of the exact data they need or are not comfortable using FTP.
GHCN Daily
GHCN (Global Historical Climatology Network)-Daily is an integrated database of daily climate summaries from land surface stations across the globe.
GHCN Monthly
Temperature dataset that contains monthly mean temperatures and is used for operational climate monitoring activities.
Comparative Climate Data
The Comparative Climatic Data tables of meteorological elements outline the climate conditions at major

- Sunshine - Average Percent of Possible (Cities Listed by Ranking Most to Least)
- Cloudiness - Mean Number of Days (Clear, Partly Cloudy, Cloudy)
- Average Relative Humidity - Morning (M), Afternoon (A)

Climate Normals

Note: The tables below contain the new 1981–2010 Normals data.

- Normal Daily Maximum Temperature, °F
- Normal Daily Minimum Temperature, °F
- Normal Daily Mean Temperature,°F
- Normal Heating Degree Days (July–June) Base 65 °F
- Normal Cooling Degree Days (January–December) Base 65 °F
- Normal Precipitation, Inches

Copy link address

Table Notes

The following gives a full explanation of all symbols and caveats used to explain the values in the above data tables.

Observed Data (Monthly and Annual)

The Observed Data values are the means and extremes for the period of record (number of years) indicated. Periods of record are documented in the Local Climatological Data annual publications.

- Temperature - Highest of Record, °F

```r
library(readr)
www = "http://www1.ncdc.noaa.gov/pub/data/ccd-data/nrmmax.txt"
temperature <- read.csv(www, header=FALSE, skip = 1)
colnames(temperature) <- c("Code","City","State","JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEP","OCT","NOV","DEC","ANN")
temperature$City <- as.character(temperature$City)

head(temperature)
```

```
##    Code          City                        State  JAN  FEB  MAR  APR
## 1 13876 BIRMINGHAM AP        AL                 30 53.8 58.4 66.7 74.4
## 2  3856    HUNTSVILLE      AL                   30 51.2 55.9 64.9 73.6
## 3 13838        MOBILE  AL                       30 60.0 63.2 69.8 76.1
## 4 13895    MONTGOMERY      AL                   30 57.4 61.8 69.7 76.6
## 5 26451     ANCHORAGE     AK                    30 23.1 26.6 33.9 44.5
## 6 25308       ANNETTE   AK                      30 41.6 42.7 44.9 50.2
##    MAY  JUN  JUL  AUG  SEP  OCT  NOV  DEC  ANN
## 1 81.5 87.7 90.8 90.6 85.1 75.3 65.4 55.9 73.8
## 2 81.3 88.2 90.7 90.9 85.0 74.6 63.7 53.5 72.8
## 3 83.0 88.2 90.4 90.5 87.3 79.4 70.3 61.9 76.7
## 4 84.0 89.8 92.1 91.9 87.3 78.3 69.0 59.6 76.5
## 5 56.0 62.8 65.4 63.5 55.1 40.5 27.8 24.8 43.7
## 6 56.3 61.1 64.3 64.7 59.3 51.6 44.6 41.5 51.9
```

```r
merged_data <- crime %>%
  filter(Year==2017) %>%
  inner_join(temperature, by = "City") %>%
  mutate(Crimepc = Violent.crime/Population)

glimpse(merged_data)
```

```
## Observations: 91
## Variables: 30
## $ State.x             <chr> "ALABAMA", "ALASKA", "ARKANSAS", "CALIFORN...
## $ City                <chr> "MONTGOMERY", "ANCHORAGE", "LITTLE ROCK", ...
## $ Year                <dbl> 2017, 2017, 2017, 2017, 2017, 2017, 2017, ...
## $ Population          <dbl> 199565, 299097, 198800, 378788, 129903, 52...
## $ Violent.crime       <dbl> 656, 1707, 1737, 929, 221, 1513, 1551, 516...
## $ Murder              <dbl> 19, 13, 28, 22, 2, 33, 11, 7, 21, 17, 0, 1...
## $ Rape                <dbl> 35, 196, 92, 48, 21, 90, 93, 25, 53, 265, ...
## $ Robbery             <dbl> 185, 369, 257, 365, 85, 502, 618, 218, 541...
## $ Aggravated.assault  <dbl> 417, 1129, 1360, 494, 113, 888, 829, 266, ...
## $ Property.crime      <dbl> 4246, 7708, 6891, 7570, 2130, 10256, 6280,...
## $ Burglary            <dbl> 1031, 1089, 1252, 1643, 203, 1860, 1442, 4...
## $ Larceny..theft      <dbl> 2797, 5180, 5038, 4621, 1499, 6908, 3416, ...
## $ Motor.vehicle.theft <dbl> 418, 1439, 601, 1306, 428, 1488, 1422, 627...
## $ Arson               <dbl> NA, 31, 29, 127, 14, 112, 59, 23, 98, 81, ...
## $ Code                <int> 13895, 26451, 13963, 23155, 14745, 93193, ...
## $ State.y             <fct>  AL                    30,  AK             ...
## $ JAN                 <dbl> 57.4, 23.1, 50.5, 56.2, 30.8, 54.8, 67.4, ...
## $ FEB                 <dbl> 61.8, 26.6, 55.1, 62.8, 34.9, 61.6, 67.2, ...
## $ MAR                 <dbl> 69.7, 33.9, 64.0, 68.7, 43.8, 67.6, 68.6, ...
## $ APR                 <dbl> 76.6, 44.5, 73.1, 75.0, 57.4, 74.6, 71.7, ...
## $ MAY                 <dbl> 84.0, 56.0, 81.1, 83.5, 68.9, 84.1, 73.6, ...
## $ JUN                 <dbl> 89.8, 62.8, 88.9, 90.9, 77.4, 92.0, 76.7, ...
## $ JUL                 <dbl> 92.1, 65.4, 92.5, 97.1, 82.3, 98.4, 81.9, ...
## $ AUG                 <dbl> 91.9, 63.5, 92.6, 95.8, 80.9, 97.1, 83.8, ...
## $ SEP                 <dbl> 87.3, 55.1, 85.6, 90.0, 72.6, 90.9, 82.1, ...
## $ OCT                 <dbl> 78.3, 40.5, 74.8, 79.4, 60.5, 79.5, 77.2, ...
## $ NOV                 <dbl> 69.0, 27.8, 63.0, 65.7, 48.4, 65.1, 72.1, ...
## $ DEC                 <dbl> 59.6, 24.8, 52.3, 56.6, 36.3, 54.9, 66.8, ...
## $ ANN                 <dbl> 76.5, 43.7, 72.8, 76.8, 57.9, 76.7, 74.1, ...
## $ Crimepc             <dbl> 0.003287150, 0.005707179, 0.008737425, 0.0...
```

```r
library(ggplot2)
# Simple scatter plot
sp <- ggplot(data=merged_data, aes(x=ANN, y=Crimepc)) + geom_point()
# adding a trend line
sp + stat_smooth(method="lm", se=FALSE)
```

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```