

# Datasets for 4e of Lock5

December 17, 2025

**Description** Documentation for the Lock5Data R package that contains datasets for the fourth edition of Statistics: Umling the Power of Data" by Lock, Lock, Lock Morgan, Lock, and Lock (Wiley 2026).

**Maintainer:** Robin Lock <rlock@stlawu.edu>

## Contents

ACS . . . . .	4
AllCountries . . . . .	5
APMultipleChoice . . . . .	6
April14Temps . . . . .	7
BaseballHits2024 . . . . .	7
BaseballSalaries2024 . . . . .	8
BaseballTimes . . . . .	9
Benford . . . . .	9
BikeCommute . . . . .	10
BMI . . . . .	11
BodyFat . . . . .	11
BodyTemp50 . . . . .	12
BootAtlantaCorr . . . . .	13
CaffeineTaps . . . . .	13
CAOSExam . . . . .	14
CarbonDioxide . . . . .	14
CarDepreciation . . . . .	15
Cars2025 . . . . .	15
CarsEV2025 . . . . .	16
Celtics2024 . . . . .	17
Cereal . . . . .	19
CityTemps . . . . .	19
CocaineTreatment . . . . .	20

ColaCalcium . . . . .	20
CollegeExperience . . . . .	21
CollegeScores . . . . .	22
CollegeScores2yr . . . . .	24
CollegeScores4yr . . . . .	25
CommuteAtlanta . . . . .	27
CommuteStLouis . . . . .	27
CompassionateRats . . . . .	28
COVIDCollege . . . . .	28
COVIDCollegeStacked . . . . .	30
CricketChirps . . . . .	31
DDS . . . . .	32
DecemberFlights . . . . .	33
DietDepression . . . . .	33
Digits . . . . .	34
DrugResistance . . . . .	35
EducationLiteracy . . . . .	35
ElectionMargin . . . . .	36
EmployedACS . . . . .	37
ExerciseHours . . . . .	37
FacebookFriends . . . . .	38
FireAnts . . . . .	39
FisherIris . . . . .	39
FishGills12 . . . . .	40
FishGills3 . . . . .	40
Flight475 . . . . .	41
FloridaLakes . . . . .	42
FootballBrain . . . . .	42
ForestFires . . . . .	43
GeneticDiversity . . . . .	44
GlobalInternet2024 . . . . .	45
GolfRound . . . . .	45
GPAbySex . . . . .	46
GroundhogDay . . . . .	46
HappyPlanetIndex . . . . .	47
HeatCognition . . . . .	48
HeightData . . . . .	48
HockeyPenalties2025 . . . . .	49
HollywoodMovies . . . . .	50
HomesForSale . . . . .	51
HomesForSaleCA . . . . .	51
HomesForSaleCanton . . . . .	52
HomesForSaleNY . . . . .	52
HomingPigeons . . . . .	53
Honeybee2023 . . . . .	54
HoneybeeCircuits . . . . .	54
HoneybeeWaggle . . . . .	55
HotDogs2024 . . . . .	55
HouseStarts2024 . . . . .	56
HumanTears25 . . . . .	56
HumanTears50 . . . . .	57
Hurricanes2024 . . . . .	58

ICUAdmissions	58
ImmuneTea	59
InkjetPrinters	60
LifeExpectancyVehicles	60
LightatNight	61
LightatNight4Weeks	62
LightatNight8Weeks	62
MalevolentUniformsNFL	63
MalevolentUniformsNHL	64
MammalLongevity	64
ManhattanApartments	65
MarriageAges	66
MastersGolf	66
MateChoice	67
MentalMuscle	67
MindsetMatters	68
MustangPrice	69
NBAPlayers2024	69
NBAStandings2024	70
NFLContracts2024	71
NFLPreseason2024	72
NFLScores2024	72
NHANES	73
Ninja	74
NutritionStudy	75
OlympicMarathon2024	76
OrganicEffect	76
OttawaSenators2024	77
PASeniors	78
PizzaGirl	79
ProcessedFoods	80
PublicLibraries	81
PumpkinBeer	81
QuizPulse10	82
RandomP50N200	82
RestaurantTips	83
RetailSales	83
RockandRoll2024	84
RollerCoasters	84
SalaryBySex	85
SampColleges	86
SampColleges2yr	87
SampColleges4yr	88
SampCountries	90
SandP500	91
SandwichAnts	92
SandwichAnts2	93
Showers	93
SkateboardPrices	94
SleepCaffeine	95
SleepStudy	95
Smiles	96

SpeedDating . . . . .	97
SplitBill . . . . .	98
StatGrades . . . . .	98
StockChanges . . . . .	99
StorySpoilers . . . . .	99
StressedMice . . . . .	100
StudentSurvey . . . . .	100
SynchronizedMovement . . . . .	101
TenCountries . . . . .	102
TextbookCosts . . . . .	103
ToenailArsenic . . . . .	103
TrafficFlow . . . . .	104
USStates . . . . .	104
WaterStriders . . . . .	106
WaterTaste . . . . .	106
Wetsuits . . . . .	107
XylitolEffect . . . . .	108
YoungBlood . . . . .	108
<b>Index</b>	<b>110</b>

ACS

*American Community Survey*

## Description

Data from a sample of individuals in the American Community Survey

## Format

A data frame with 10,000 observations on the following 9 variables.

Sex 0=female and 1=male

Age Age (years)

Married 0=not married and 1=married

Income Wages and salary for the past 12 months (in \$1,000's)

HoursWk Hours of work per week

Race Asian, Black, Other, or White

USCitizen 1=citizen and 0=noncitizen

HealthInsurance 1=have health insurance and 0=no health insurance

Language 1=English spoken at home and 0=other

## Details

The American Community Survey, administered by the US Census Bureau, is given every year to a random sample of about 3.5 million households (about 3% of all US households). Data on a random sample of 1% of all US residents are made public (after ensuring anonymity), and we have selected a random sub-sample of  $n = 10000$  from the 2023 data for this dataset.

\*\* Updated for 4e (earlier versions are ACS3e (from 2017) and ACS2010). \*\*

**Source**

The full public dataset can be downloaded at <https://www.census.gov/programs-surveys/acs/microdata.html>, and the full list of variables are at <https://www.census.gov/programs-surveys/acs/microdata/documentation.html>

Specific request for these variables is <https://data.census.gov/app/mdat/ACSPUMS1Y2023/table?cv=SEX,MAR,RAC1P,HICOV&vv=AGEP,WAGP,WKHP&rv=CIT,LANX&wt=PWGTP>

---

AllCountries	<i>All Countries</i>
--------------	----------------------

---

**Description**

Data on the countries of the world

**Format**

A data frame with 217 observations on the following 29 variables.

Country Country name  
 Code Three-letter code for country  
 LandArea Size in 1000 sq. km.  
 Population Population in millions  
 Density Number of people per square kilometer  
 GDP Gross Domestic Product (in \$US) per capita  
 Rural1 Percentage of population living in rural areas  
 CO2 CO2 emissions (metric tons per capita)  
 PumpPrice Price for a liter of gasoline (\$US)  
 Military Percentage of government expenditures directed toward the military  
 Health Percentage of government expenditures directed towards healthcare  
 ArmedForces Number of active duty military personnel (in 1,000's)  
 Internet Percentage of the population with access to the internet  
 Cell Cell phone subscriptions (per 100 people)  
 HIV Percentage of the population with HIV  
 Hunger Percent of the population considered undernourished  
 Diabetes Percent of the population diagnosed with diabetes  
 BirthRate Births per 1000 people  
 DeathRate Deaths per 1000 people  
 ElderlyPop Percentage of the population at least 65 years old  
 LifeExpectancy Average life expectancy (years)  
 FemaleLabor Percent of females 15 - 64 in the labor force  
 Unemployment Percent of labor force unemployed  
 Renewable Percent of energy from renewable sources

Energy Total energy consumption (million BTU per capita)

Electricity Electric power consumption (kWh per capita)

Developed Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000

HDI Human Development Index - United Nations' measure of social and economic well being on a 0-1 scale

HDIGroup Categories (Very High, High, Medium, Low) based on HDI

### Details

Most data for each variable were collected for 2023 (or most recently available year) from <https://data.worldbank.org/>. Energy and Electricity values come from U.S. Energy Information Administration. HDI values from United Nations Human Development Report.

\*\* This dataset is updated for 4e from earlier versions (now Allcountries1e, AllCountries2e, and All Countries3e) \*\*

### Source

Most data were gathered online from <https://data.worldbank.org/>.

Gasoline prices come from <https://tradingeconomics.com/country-list/gasoline-prices?continent=world>.

Electricity and Energy variables from U.S. Energy Information Administration, <https://www.eia.gov/international/data/world#/>

HDI variables from United Nations Human Development Report, <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI> All accessed January 2025.

---

APMultipleChoice

*AP Multiple Choice*

---

### Description

Correct responses on Advanced Placement multiple choice exams

### Format

A dataset with 400 observations on the following variable.

Answer Correct response: A, B, C, D, or E

### Details

Correct responses from multiple choice sections for a sample of released Advanced Placement exams

### Source

Sample exams from several disciplines at <http://apcentral.collegeboard.com>

---

April14Temps*April 14th Temperatures*

---

**Description**

Temperatures in Des Moines, IA and San Francisco, CA on April 14th

**Format**

A data frame with 30 observations on the following 3 variables.

Year 1995 to 2024

DesMoines Temperature in Des Moines (degrees F)

SanFrancisco Temperature in San Francisco (degrees F)

**Details**

Average temperature for the day of April 14th in each of 30 years from 1995-2024

\*\* Data set updated for 4e (earlier versions are now April14Temps3e, April14Temps2e, and April14Temps1e)  
\*\*

**Source**

The University of Dayton Average Daily Temperature Archive at <https://academic.udayton.edu/kissock/http/Weather/citylistUS.htm>

Recent updates from <https://www.wunderground.com/history/daily/us/ca/san-francisco/KSF0> and <https://www.wunderground.com/history/daily/us/ia/des-moines/KDSM>

---

BaseballHits2024*Baseball Team Statistics (2024)*

---

**Description**

Number of hits, wins, and other stats for MLB teams in 2024

**Format**

A data frame with 30 observations on the following 14 variables.

Team Name of baseball team (3-character code)

League Either AL or NL

Wins Number of wins for the season

Runs Number of runs scored

Hits Number of hits

Doubles Number of doubles

Triples Number of triples

HomeRuns Number of home runs  
RBI Number of runs batted in  
StolenBases Number of stolen bases  
CaughtStealing Number of times caught stealing  
Walks Number of walks  
Strikeouts Number of strikeouts  
BattingAvg Team batting average

**Details**

Offensive team statistics for the 2024 Major League Baseball regular season.  
\*\* Updated for 4e (earlier versions are now BaseballHits2019, BaseballHits2014, and Baseball-Hits1e)

**Source**

<https://www.baseball-reference.com/leagues/MLB/2024-standard-batting.shtml>

---

BaseballSalaries2024    *MLB Player Salaries in 2024*

---

**Description**

Opening Day salaries for all Major League Baseball players in 2024

**Format**

A data frame with 952 observations on the following 4 variables.  
  
Name Player's name  
Salary 2024 season salary (in millions)  
Team Team name  
Pos Code for player's main position

**Details**

Yearly salary (in millions of dollars) for all players on the rosters of Major League Baseball teams at the start of the 2024 season.  
\*\* Updated for 4e (earlier versions are BaseballSalaries2019 and BaseballSalaries2015). \*\*

**Source**

Downloaded from USA Today at <https://databases.usatoday.com/major-league-baseball-salaries-2024/>. (March 2025)



---

BaseballTimes*Baseball Game Times (2024)*

---

**Description**

Information for a sample of 50 Major League Baseball games played during the 2024 season

**Format**

A dataset with 50 observations on the following 9 variables.

Away	Away team name
Home	Home team name
Runs	Total runs scored (both teams)
Margin	Margin of victory
Hits	Total number of hits (both teams)
Errors	Total number of errors (both teams)
Pitchers	Total number of pitchers used (both teams)
Walks	Total number of walks (both teams)
Time	Elapsed time for game (in minutes)

**Details**

Data from a sample of boxscores for Major League Baseball games played during the 2024 season. Games include all played on May 6th, June 6th, July 13th, and August 16th,

**Source**

Data obtained from boxscores at <https://www.baseball-reference.com/boxes/index.fcgi>

---

Benford*Benford data*

---

**Description**

Two examples to test Benford's Law

**Format**

A dataset with 9 observations on the following 4 variables.

Digit	Leading digit (1-9)
BenfordP	Expected proportion according to Benford's law
Address	Frequency as a first digit in an address
Invoices	Frequency as the first digit in invoice amounts

**Details**

Leading digits from 1188 addresses sampled from a phone book and 7273 amounts from invoices sampled at a company.

**Source**

Thanks to Prof. Richard Cleary for providing the data

---

BikeCommute	<i>Bike Commute</i>
-------------	---------------------

---

**Description**

Commute times for two kinds of bicycle

**Format**

A dataset with 56 observations on the following 9 variables.

Bike	Type of material Carbon or Steel
Date	Date of the bike commute
Distance	Length of commute (in miles)
Time	Total commute time (hours:minutes:seconds)
Minutes	Time converted to minutes
AvgSpeed	Average speed during the ride (miles per hour)
TopSpeed	Maximum speed (miles per hour)
Seconds	Time converted to seconds
Month	Categories: 1Jan 2Feb 3Mar 4Apr 5May 6June 7July

**Details**

Data from a personal experiment to compare commuting time based on a randomized selection between two bicycles made of different materials.

**Source**

Thanks to Dr. Groves for providing his data.

**References**

Bicycle weight and commuting time: randomised trial, in British Medical Journal, BMJ 2010;341:c6801.

BMI

*Body Mass Index and Exercise***Description**

Body Mass Index (BMI) and exercise indicator from the 2023 BRFSS survey

**Format**

A data frame with 392,788 observations on the following 2 variables.

BMI Body Mass Index = weight (in kg) / (height (in m))^2

Exercise Have you exercised in last 30 days? (Yes or No)

**Details**

Data from the 2023 Behavioral Risk Factor Surveillance System (BRFSS) survey. BMI is calculated as weight (in kg) divided by height (in meters) squared. Values above 25 are considered overweight and over 30 are classified as obese.

**Source**

Centers for Disease Control and Prevention (CDC), Behavioral Risk Factor Surveillance System Survey Data, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta, 2023. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2023.html](https://www.cdc.gov/brfss/annual_data/annual_2023.html)

BodyFat

*Body Measurements***Description**

Percent fat and other body measurements for a sample of men

**Format**

A dataset with 100 observations on the following 10 variables.

Bodyfat	Percent body fat
Age	Age in years
Weight	Weight in pounds
Height	Height in inches
Neck	Neck circumference in cm.
Chest	Chest circumference in cm.
Abdomen	Abdomen circumference in cm.
Ankle	Ankle circumference in cm.
Biceps	Extended biceps circumference in cm.
Wrist	Wrist circumference in cm.

**Details**

This is a subset of a larger sample of men who each had a percent body fat estimated by an underwater weighing technique. Other measurements were taken to see how they might be used to predict the body fat percentage.

**Source**

These data were contributed by Roger Johnson, then at Carleton University, to the Datasets Archive at the Journal of Statistics Education.

<https://www2.amstat.org/publications/jse/v4n1/datasets.johnson.html>

The data were originally supplied by Dr. A. Garth Fisher, Human Performance Research Center, Brigham Young University, Provo, Utah 84602.

---

BodyTemp50

*Body Temperatures*

---

**Description**

Sample of 50 body temperatures

**Format**

A data frame with 50 observations on the following 3 variables.

BodyTemp Body temperature in degrees F

Pulse Pulse rates (beat per minute)

Sex F=Female, M=Male

**Details**

Body temperatures and pulse rates for a sample of 50 healthy adults. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.

**Source**

Shoemaker, "What's Normal: Temperature, Gender and Heartrate", Journal of Statistics Education, Vol. 4, No. 2 (1996)

<http://jse.amstat.org/v4n2/datasets.shoemaker.html>

---

 BootAtlantaCorr

*Bootstrap Correlations for Atlanta Commutes*


---

**Description**

Bootstrap correlations between Time and Distance for 500 commuters in Atlanta

**Format**

A dataset with 1000 observations on the following variable.

CorrTimeDist    Correlation between Time and Distance for a bootstrap sample of Atlanta commuters

**Details**

Correlations for bootstrap samples of Time vs. Distance for the data on Atlanta commuters in CommuteAtlanta.

**Source**

Computer simulation

---

 CaffeineTaps

*Caffeine Taps*


---

**Description**

Finger tap rates with and without caffeine

**Format**

A dataset with 20 observations on the following 2 variables.

Taps	Number of finger taps in one minute
Group	Treatment with levels Caffeine NoCaffeine

**Details**

Results from a double-blind experiment where a sample of male college students were asked to tap their fingers at a rapid rate. The sample was then divided at random into two groups of ten students each. Each student drank the equivalent of about two cups of coffee, which included about 200 mg of caffeine for the students in one group but was decaffeinated coffee for the second group. After a two hour period, each student was tested to measure finger tapping rate (taps per minute). The goal of the experiment was to determine whether caffeine produces an increase in the average tap rate.

**Source**

Hand, Daly, Lund, McConway and Ostrowski, Handbook of Small Data Sets, Chapman and Hall, London (1994), pp. 40

CAOSExam

*CAOS Exam Scores***Description**

Scores on a pre-test and post-test of basic statistics concepts

**Format**

A dataset with 10 observations on the following 3 variables.

Student	ID code for student
Pretest	CAOS Pretest score
Posttest	CAOS Posttest score

**Details**

The CAOS (Comprehensive Assessment of Outcomes in First Statistics Course) exam is designed to measure comprehension of basic statistical ideas in an introductory statistics course. This dataset has scores for ten students who took the CAOS pre-test at the start of a course and the post-test during the course itself. Each exam consists of 40 multiple choice questions and the score is the percentage correct.

**Source**

A sample of 10 students from an introductory statistics course. Find out more about the CAOS exam at <http://app.gen.umn.edu/artist/caos.html>

CarbonDioxide

*Carbon Dioxide Levels***Description**

Atmospheric carbon dioxide levels by year

**Format**

A data frame with 13 observations on the following 2 variables.

Year	Every five years from 1960 to 2020
CO2	Carbon dioxide level in parts per million

**Details**

Carbon dioxide levels in the atmosphere over a 60 year span from 1960-2020.

\*\* Updated for 4e (earlier versions are now CarbonDioxide3e and CarbonDioxide2e) \*\*

**Source**

Dr. Pieter Tans, NOAA/ESRL. Values recorded at the Mauna Loa Observatory in Hawaii. <https://gml.noaa.gov/ccgg/trends/>

---

CarDepreciation	<i>Car Depreciation</i>
-----------------	-------------------------

---

**Description**

Depreciation for 20 car models (2024).

**Format**

A dataset with 20 observations on the following 5 variables.

Car	Name of the car model
New	Price of a new car
Used	Value after one year
Depreciation	Drop in value after one year
PctDrop	Percent depreciation after one year

**Details**

Twenty car models were selected at random. Original price (in dollars) and value after one year (and 12,000 miles) were recorded for each model. The depreciation is the difference (New-Used). Updated for 4e (earlier version is now CarDepreciation3e)

**Source**

New and used costs determined using models selected from <https://caredge.com/depreciation>.

---

Cars2025	<i>2025 Car Models</i>
----------	------------------------

---

**Description**

Information about new car models in 2025

**Format**

A data frame with 107 observations on the following 21 variables.

Make	Manufacturer (e.g. Chevrolet, Toyota, etc.)
Model	Car model (e.g. Impala, Highlander, ...)
Type	Vehicle category (Pickup, Hatchback, Minivan, Sedan, Sporty, SUV, or Wagon)
LowPrice	Lowest MSRP (in \$1,000)
HighPrice	Highest MSRP (in \$1,000)
CityMPG	City miles per gallon (EPA)
HwyMPG	Highway miles per gallon (EPA)
Seating	Seating capacity
Drive	Type of drive (AWD, FWD, or RWD)

Acc030 Time (in seconds) to go from 0 to 30 mph  
 Acc060 Time (in seconds) to go from 0 to 60 mph  
 QtrMile Time (in seconds) to go ¼ mile  
 Braking Distance to stop from 60 mph (dry pavement)  
 FuelCap Fuel capacity (in gallons)  
 Length Length (in inches)  
 Width Width (in inches)  
 Height Height (in inches)  
 Wheelbase Wheelbase (in inches)  
 UTurn Diameter (in feet) needed for a U-turn  
 Weight Curb weight (in pounds)  
 Size Large, Midsized, or Small

### Details

Data for a set of 107 new car models in 2025 based on information in Consumer Reports.

\*\* Updated for 4e (an earlier versions sre Cars2020 and Cars2015). \*\*

### Source

Data on new car models in 2025 accessed from Consumer Reports website. <https://www.consumerreports.org/cars/>

---

CarsEV2025	<i>2025 EV Car Models</i>
------------	---------------------------

---

### Description

Information about new electric, plug-in hybrid, and hybrid car models in 2025

### Format

A data frame with 93 observations on the following 25 variables.

Make Manufacturer (e.g. Chevrolet, Toyota, etc.)  
 Model Car model (e.g. Camry Hybrid, Blazer EV, ...)  
 Type Vehicle category (Pickup, Hatchback, Minivan, Sedan, SUV, or Wagon)  
 Class BEV= all electric, PHEV=plug-in hybrid, Hybrid=hybrid  
 LowPrice Lowest MSRP (in \$1,000)  
 HighPrice Highest MSRP (in \$1,000)  
 Seating Seating capacity  
 Acc030 Time (in seconds) to go from 0 to 30 mph  
 Acc060 Time (in seconds) to go from 0 to 60 mph  
 QtrMile Time (in seconds) to go ¼ mile  
 Braking Distance to stop from 60 mph (dry pavement)



MPKWH Miles per Kilowatt hour of electricity  
 MPGE Miles per gallon of gas equivalent  
 MPG Miles per gallon when using gas  
 RangeE Range when using battery only  
 RangeG Range when using gas  
 Battery Size of battery (in kWh)  
 FuelCap Gas fuel capacity (in gallons)  
 Length Length (in inches)  
 Width Width (in inches)  
 Height Height (in inches)  
 Wheelbase Wheelbase (in inches)  
 UTurn Diameter (in feet) needed for a U-turn  
 Weight Curb weight (in pounds)  
 CRType Consumer report classification: Car, Luxury, Luxury SUV, Minivan, Pickup, SUV, SUV-3Row

### Details

Data for a set of 93 new electric, plug-in hybride, or hybrid car models in 2025 based on information in Consumer Reports.

### Source

Data on new car models in 2025 accessed from Consumer Reports website. <https://www.consumerreports.org/cars/>

---

Celtics2024

*Boston Celtics Basketball (2024)*

---

### Description

Game log data for the Boston Celticss basketball team in 2023-2024 regular season

### Format

A data frame with 82 observations on the following 33 variables.

Game ID number for each game  
 Date Date the game was played (mm/dd/yyyy)  
 Location Away or Home  
 Opp Opponent team  
 Win Game result: L or W  
 Points Number of points scored  
 FG Field goals made  
 FGA Field goals attempted

FG3 Three-point field goals made  
 FG3A Three-point field goals attempted  
 FT Free throws made  
 FTA Free throws attempted  
 Rebounds Total rebounds  
 OffReb Offensive rebounds  
 Assists Number of assists  
 Steals Number of steals  
 Blocks Number of shots blocked  
 Turnovers Number of turnovers  
 Fouls Number of fouls  
 OppPoints Opponent's points scored  
 OppFG Opponent's field goals made  
 OppFGA Opponent's field goals attempted  
 OppFG3 Opponent's three-point field goals made  
 OppFG3A Opponent's three-point field goals attempted  
 OppFT Opponent's free throws made  
 OppFTA Opponent's free throws attempted  
 OppRebounds Opponent's total rebounds  
 OppOffReb Opponent's offensive rebounds  
 OppAssists Opponent's assists  
 OppSteals Opponent's steals  
 OppBlocks Opponent's shots blocked  
 OppTurnovers Opponent's turnovers  
 OppFouls Opponent's fouls

### Details

Information from online boxscores for all 82 regular season games played by the Boston Celtics basketball team during the 2023-2024 season.

\*\* Updated for 4e (earlier versions for the Golden State Warriors are GSWarriors2019 and GSWarriors2016. The 1e version is MiamiHeat dataset from 2011.) \*\*

### Source

Data for the 2023-2024 Boston Celtics games downloaded from <https://www.basketball-reference.com/teams/BOS/2024/gamelog/> (March 2025)

---

Cereal	<i>Breakfast Cereals</i>
--------	--------------------------

---

**Description**

Nutrition information for a sample of 30 breakfast cereals

**Format**

A dataset with 30 observations on the following 10 variables.

Name	Brand name of cereal
Company	Manufacturer coded as G=General Mills, K=Kellogg's or Q=Quaker
Serving	Serving size (in cups)
Calories	Calories (per cup)
Fat	Fat (grams per cup)
Sodium	Sodium (mg per cup)
Carbs	Carbohydrates (grams per cup)
Fiber	Dietary Fiber (grams per cup)
Sugars	Sugars (grams per cup)
Protein	Protein (grams per cup)

**Details**

Nutrition contents for a sample of breakfast cereals, derived from nutrition labels. Values are per cup of cereal (rather than per serving).

**Source**

Cereal data obtained from nutrition labels at  
<http://www.nutritionresource.com/foodcomp2.cfm?id=0800>

---

CityTemps	<i>City Temperatures</i>
-----------	--------------------------

---

**Description**

Mean monthly temperature in Moscow, Melbourne, and San Francisco for 2023 and 2024

**Format**

A data frame with 24 observations on the following 5 variables.

Year 2023 or 2024

Month 1=January through 12=December

Moscow Monthly temperatures in Moscow (Russia)

Melbourne Monthly temperatures in Melbourne (Australia)

SanFrancisco Monthly temperatures in San Francisco (United States)

**Details**

Mean monthly temperatures in degrees C for the years 2023 and 2024 in each of three cities.

\*\* Updated for 4e (earlier versions for 2017 and 2018 in CityTemps3e, for 2014 and 2015 in CityTemps2e). \*\*

**Source**

Source: <https://chhttps://www.weatherandclimate.info/history/> add station codes 94866 (Melbourne), 72494 (San Francisco), 27612 (Moscow) to url.

---

CocaineTreatment

*Cocaine Treatment*

---

**Description**

Relapse/no relapse responses to three different treatments for cocaine addiction

**Format**

A dataset with 72 observations on the following 2 variables.

Drug	Treatment drug: Desipramine, Lithium, or Placebo
Relapse	Did the patient relapse? no or yes

**Details**

Data from an experiment to investigate the effectiveness of the two drugs, desipramine and lithium, in the treatment of cocaine addiction. Subjects (cocaine addicts seeking treatment) were randomly assigned to take one of the treatment drugs or a placebo. The response variable is whether or not the subject relapsed (went back to using cocaine) after the treatment.

**Source**

Gawin, F., et.al., "Desipramine Facilitation of Initial Cocaine Abstinence", Archives of General Psychiatry, 1989; 46(2): 117 - 121.

---

ColaCalcium

*Cola Calcium*

---

**Description**

Calcium excretion with diet cola and water

**Format**

A dataset with 16 observations on the following 2 variables.

Drink	Type of drink: Diet cola or Water
Calcium	Amount of calcium excreted (in mg.)

**Details**

A sample of 16 healthy women aged 18 - 40 were randomly assigned to drink 24 ounces of either diet cola or water. Their urine was collected for three hours after ingestion of the beverage and calcium excretion (in mg.) was measured. The researchers were investigating whether diet cola leaches calcium out of the system, which would increase the amount of calcium in the urine for diet cola drinkers.

**Source**

Larson, Amin, Olsen, and Poth, Effect of Diet Cola on Urine Calcium Excretion, *Endocrine Reviews*, 31[3]: S1070, June 2010. These data are recreated from the published summary statistics, and are estimates of the actual data.

---

CollegeExperience	<i>College Experience (2021-2022)</i>
-------------------	---------------------------------------

---

**Description**

Measures on college students in the 2021-2022 school year

**Format**

A data frame with 79 observations on the following 13 variables. Values are averages over reports throughout the year for each student.

**iOS** Indicator for whether the student is an iOS user (1) or an Android user (0)

**Still** Time spent still (no movement) in seconds

**Steps** Number of steps per day

**PhoneUnlocked** Amount of time that a student uses their phone in unlocked state (i.e., total phone usage duration)

**PhoneProportion** Proportion of waking hours that a student spends with their phone unlocked

**Sleep** Sleep duration (based on a predictive model) in hours

**Anxious** Over the last 2 weeks, how often have you been bothered by the following problems?  
Feeling nervous, anxious or on edge (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

**Depressed** Over the last 2 weeks, how often have you been bothered by the following problems?  
Feeling down, depressed or hopeless (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

**Social** Have you spent most of your time alone or with others today? (1: Almost always alone; 2: Mostly alone, a little time with others; 3: Equal amounts of time alone and with others; 4: Mostly with others, a little time alone; 5: Almost always with others)

**FeelGood** Right now, Overall, I feel good about myself (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

**Stress** Are you feeling stressed now? (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

**Sex** Self-reported sex of the participant (F or M)

**AnxiousBinary** 1 if the value for Anxious is > 0, 0 if the value is 0.

### Details

College students allowed researchers to collect live data readily accessible from their smartphones (such as data on phone usage, physical activity, and sleep) for the duration of their time at college. These students also regularly assessed their own mental health and well-being via surveys. This dataset contains each student's average values for the 2021-2022 academic year.

### Source

Subigya Nepal, et. al. "Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience, and Behavior of College Students during the Pandemic", Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1, Article 38 (March 2024). <https://dl.acm.org/doi/10.1145/3643501>.

---

CollegeScores

*College Scorecard*

---

### Description

Information on all US post-secondary schools collected by the Department of Education for the College Scorecard

### Format

A data frame with 5702 observations on the following 35 variables.

**Name** Name of the school

**State** State where school is located

**ID** ID number for school

**Main** Main campus? (1=yes, 0=branch campus)

**Accred** Accreditation agency

**MainDegree** Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)

**HighDegree** Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)

**Control** Control of school (Private, Profit, Public)

**Region** Region of country (Midwest, Northeast, Southeast, Territory, West)

**Locale** Locale (City, Rural, Suburb, Town)

**Latitude** Latitude

**Longitude** Longitude

**AdmitRate** Admission rate

MidACT	Median of ACT scores
AvgSAT	Average combined SAT equivalent scores for admitted students
Online	Only online (distance) programs
Enrollment	Undergraduate enrollment
White	Percent of undergraduates who report being white
Black	Percent of undergraduates who report being black
Hispanic	Percent of undergraduates who report being Hispanic
Asian	Percent of undergraduates who report being Asian
Other	Percent of undergraduates who don't report one of the above
PartTime	Percent of undergraduates who are part-time students
NetPrice	Average net price (cost minus aid)
Cost	Average total cost for tuition, room, board, etc.
TuitionIn	In-state tuition and fees
TuitionOut	Out-of-state tuition and fees
TuitionFTE	Net Tuition revenue per FTE student
InstructFTE	Instructional spending per FTE student
FacSalary	Average monthly salary for full-time faculty
FullTimeFac	Percent of faculty that are full-time
Pell	Percent of students receiving Pell grants
CompRate	Completion rate (percent who finish program within 150% of normal time)
Debt	Median debt for students who complete program
PctWomen	Percent of women students

## Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subset of the variables in the full College Scorecard. Schools with missing or zero enrollment were omitted. Updated for 4e (previous dataset is now CollegeScores3e).

## Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

CollegeScores2yr

*College Scorecard - Two Year***Description**

Information on all US colleges and universities that primarily grant associate's degrees, collected by the Department of Education for the College Scoreboard.

**Format**

A data frame with 1037 observations on the following 35 variables.

Name Name of the school

State State where school is located

ID ID number for school

Main Main campus? (1=yes, 0=branch campus)

Accred Accreditation agency

MainDegree Predominant undergrad degree (2=associate)

HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)

Control Control of school (Private, Profit, Public)

Region Region of country (Midwest, Northeast, Southeast, Territory, West)

Locale Locale (City, Rural, Suburb, Town)

Latitude Latitude

Longitude Longitude

AdmitRate Admission rate

MidACT Median of ACT scores

AvgSAT Average combined SAT equivalent scores for admitted students

Online Only online (distance) programs

Enrollment Undergraduate enrollment

White Percent of undergraduates who report being white

Black Percent of undergraduates who report being black

Hispanic Percent of undergraduates who report being Hispanic

Asian Percent of undergraduates who report being Asian

Other Percent of undergraduates who don't report one of the above

PartTime Percent of undergraduates who are part-time students

NetPrice Average net price (cost minus aid)

Cost Average total cost for tuition, room, board, etc.

TuitionIn In-state tuition and fees

TuitionOut Out-of-state tuition and fees

TuitionFTE Net Tuition revenue per FTE student

InstructFTE Instructional spending per FTE student

FacSalary Average monthly salary for full-time faculty



FullTimeFac Percent of faculty that are full-time  
 Pell Percent of students receiving Pell grants  
 CompRate Completion rate (percent who finish program within 150% of normal time)  
 Debt Median debt for students who complete program  
 PctWomen Percent of women students

### Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subset of the variables in the full College Scorecard and only the schools that primarily grant associate's degrees (MainDegree=2). The CollegeScores dataset contains these and other schools with other degree types. Schools with missing or zero enrollment were omitted. Updated for 4e (previous dataset is now CollegeScores2yr3e).

### Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

---

CollegeScores4yr	<i>College Scorecard - Four Year</i>
------------------	--------------------------------------

---

### Description

Information on all US colleges and universities that primarily grant bachelor's degrees, collected by the Department of Education for the College Scoreboard

### Format

A data frame with 2007 observations on the following 35 variables.

Name Name of the school  
 State State where school is located  
 ID ID number for school  
 Main Main campus? (1=yes, 0=branch campus)  
 Accred Accreditation agency  
 MainDegree Predominant undergrad degree (3=bachelors)  
 HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)  
 Control Control of school (Private, Profit, Public)  
 Region Region of country (Midwest, Northeast, Southeast, Territory, West)  
 Locale Locale (City, Rural, Suburb, Town)  
 Latitude Latitude  
 Longitude Longitude  
 AdmitRate Admission rate  
 MidACT Median of ACT scores

AvgSAT Average combined SAT equivalent scores for admitted students  
 Online Only online (distance) programs  
 Enrollment Undergraduate enrollment  
 White Percent of undergraduates who report being white  
 Black Percent of undergraduates who report being black  
 Hispanic Percent of undergraduates who report being Hispanic  
 Asian Percent of undergraduates who report being Asian  
 Other Percent of undergraduates who don't report one of the above  
 PartTime Percent of undergraduates who are part-time students  
 NetPrice Average net price (cost minus aid)  
 Cost Average total cost for tuition, room, board, etc.  
 TuitionIn In-state tuition and fees  
 TuitionOut Out-of-state tuition and fees  
 TuitionFTE Net Tuition revenue per FTE student  
 InstructFTE Instructional spending per FTE student  
 FacSalary Average monthly salary for full-time faculty  
 FullTimeFac Percent of faculty that are full-time  
 Pell Percent of students receiving Pell grants  
 CompRate Completion rate (percent who finish program within 150% of normal time)  
 Debt Median debt for students who complete program  
 PctWomen Percent of women students

## Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains a small subset of the variables in the full College Scorecard and only the schools that primarily grant bachelor's degrees (MainDegree=3). The CollegeScores dataset contains these and other schools with other degree types. Schools with missing or zero enrollment were omitted. Updated for 4e (previous dataset is now CollegeScores4yr3e).

## Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

---

CommuteAtlanta*Commute Atlanta*

---

**Description**

Commute times and distances for a sample of 500 people in Atlanta

**Format**

A data frame with 500 observations on the following 5 variables.

City	Atlanta
Age	Age of the respondent (in years)
Distance	Commute distance (in miles)
Time	Commute time (in minutes)
Sex	F or M

**Details**

Data from the US Census Bureau's American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted from respondents in the Atlanta metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

**Source**

Sample chosen using DataFerret at <http://www.thedataweb.org/index.html>.

---

CommuteStLouis*Commute Times in St. Louis*

---

**Description**

Commute times and distances for a sample of 500 people in St. Louis

**Format**

A dataset with 500 observations on the following 5 variables.

City	St. Louis
Age	Age of the respondent (in years)
Distance	Commute distance (in miles)
Time	Commute time (in minutes)
Sex	F or M

**Details**

Data from the US Census Bureau's American Housing Survey (AHS) which contains information about housing and living conditions for samples from certain metropolitan areas. These data were extracted from respondents in the St. Louis metropolitan area. They include only cases where the respondent worked somewhere other than home. Values show the time (in minutes) and distance (in miles) that respondents typically traveled on their commute to work each day as well as age and sex.

**Source**

Sample chosen using DataFerret at <http://www.thedataweb.org/index.html>.

---

CompassionateRats	<i>Compassionate Rats</i>
-------------------	---------------------------

---

**Description**

Would a rat attempt to free a trapped rat?

**Format**

A dataset with 30 observations on the following 2 variables.

Sex	Sex of the rat: coded as F or M
Empathy	Freed the trapped rat? no or yes

**Details**

In a recent study, some rats showed compassion by freeing another trapped rat, even when chocolate served as a distraction and even when the rats would then have to share the chocolate with their freed companion.

**Source**

Bartal I.B., Decety J., and Mason P., "Empathy and Pro-Social Behavior in Rats," Science, 2011; 224(6061):1427-1430.

---

COVIDCollege	<i>College Experience Before, During, After Covid</i>
--------------	---

---

**Description**

Measures on college students before, during, and after the Covid-19 pandemic

**Format**

A data frame with 188 observations on the following 28 variables. Values are averages over days in three different period: Before=prior to March 2020, During=March 2020 to July 2021, After = after July 2021.

StillBefore Time spent still (no movement) in seconds

StepsBefore Number of steps per day

PhoneUnlockedBefore Amount of time that a student uses their phone in unlocked state (i.e., total phone usage duration)

SleepBefore Sleep duration (based on a predictive model) in hours

AnxiousBefore Over the last 2 weeks, how often have you been bothered by the following problems? Feeling nervous, anxious or on edge (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

DepressedBefore Over the last 2 weeks, how often have you been bothered by the following problems? Feeling down, depressed or hopeless (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

SocialBefore Have you spent most of your time alone or with others today? (1: Almost always alone; 2: Mostly alone, a little time with others; 3: Equal amounts of time alone and with others; 4: Mostly with others, a little time alone; 5: Almost always with others)

FeelGoodBefore Right now, Overall, I feel good about myself (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

StressBefore Are you feeling stressed now? (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

Sex Self-reported sex of the participant (F, M, or both)

StillDuring Time spent still (no movement) in seconds

StepsDuring Number of steps per day

PhoneUnlockedDuring Amount of time that a student uses their phone in unlocked state (i.e., total phone usage duration)

SleepDuring Sleep duration (based on a predictive model) in hours

AnxiousDuring Over the last 2 weeks, how often have you been bothered by the following problems? Feeling nervous, anxious or on edge (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

DepressedDuring Over the last 2 weeks, how often have you been bothered by the following problems? Feeling down, depressed or hopeless (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

SocialDuring Have you spent most of your time alone or with others today? (1: Almost always alone; 2: Mostly alone, a little time with others; 3: Equal amounts of time alone and with others; 4: Mostly with others, a little time alone; 5: Almost always with others)

FeelGoodDuring Right now, Overall, I feel good about myself (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

StressDuring Are you feeling stressed now? (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

StillAfter Time spent still (no movement) in seconds

StepsAfter Number of steps per day

PhoneUnlockedAfter Amount of time that a student uses their phone in unlocked state (i.e., total phone usage duration)

SleepAfter Sleep duration (based on a predictive model) in hours

AnxiousAfter Over the last 2 weeks, how often have you been bothered by the following problems? Feeling nervous, anxious or on edge (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

DepressedAfter Over the last 2 weeks, how often have you been bothered by the following problems? Feeling down, depressed or hopeless (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

SocialAfter Have you spent most of your time alone or with others today? (1: Almost always alone; 2: Mostly alone, a little time with others; 3: Equal amounts of time alone and with others; 4: Mostly with others, a little time alone; 5: Almost always with others)

FeelGoodAfter Right now, Overall, I feel good about myself (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

StressAfter Are you feeling stressed now? (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

### Details

The COVID-19 pandemic changed college life dramatically for many students. Like many colleges, the college where these data were collected suddenly went remote in March 2020, and for the 2020-2021 school year only about half of the students attended in-person, most classes were held online, and there were strict restrictions on social gatherings. Campus operations returned mostly to normal for the 2021-2022 school year. Students in college from 2018-2022 spanned college life before, during, and after the pandemic-induced changes to college life, and this dataset includes various mobile sensing and mental health variables averaged over these three time periods. \

Note: The dataset **CollegeExperience** includes many of the same variables on many of the same students, but is limited to data from the 2021-2022 academic year.

### Source

Subigya Nepal, et. al. "Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience, and Behavior of College Students during the Pandemic", Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1, Article 38 (March 2024). <https://dl.acm.org/doi/10.1145/3643501>.

---

COVIDCollegeStacked      *College Experience Before, During, After Covid (Stacked Format)*

---

### Description

Measures on college students before, during, and after the Covid-19 pandemic

### Format

A data frame with 492 observations on the following 11 variables. Values are averages over days in each time period.

**Period** Before=prior to March 2020, During=March 2020 to July 2021, After = after July 2021

Still Time spent still (no movement) in seconds

Steps Number of steps per day

**PhoneUnlocked** Amount of time that a student uses their phone in unlocked state (i.e., total phone usage duration)

**Sleep** Sleep duration (based on a predictive model) in hours

**Anxious** Over the last 2 weeks, how often have you been bothered by the following problems?  
Feeling nervous, anxious or on edge (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

**Depressed** Over the last 2 weeks, how often have you been bothered by the following problems?  
Feeling down, depressed or hopeless (0: Not at all; 1: Several days; 2: More than half the days; 3: Nearly every day)

**Social** Have you spent most of your time alone or with others today? (1: Almost always alone; 2: Mostly alone, a little time with others; 3: Equal amounts of time alone and with others; 4: Mostly with others, a little time alone; 5: Almost always with others)

**FeelGood** Right now, Overall, I feel good about myself (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

**Stress** Are you feeling stressed now? (1: Not at All; 2: A Little Bit; 3: Somewhat; 4: Very Much; 5: Extremely)

**Sex** Self-reported sex of the participant (F, M, or both)

## Details

The COVID-19 pandemic changed college life dramatically for many students. Like many colleges, the college where these data were collected suddenly went remote in March 2020, and for the 2020-2021 school year only about half of the students attended in-person, most classes were held online, and there were strict restrictions on social gatherings. Campus operations returned mostly to normal for the 2021-2022 school year. Students in college from 2018-2022 spanned college life before, during, and after the pandemic-induced changes to college life, and this dataset includes various mobile sensing and mental health variables averaged over these three time periods.

Note: This dataset, **COVIDCollegeStacked**, contains the same data as **COVIDCollege**, but rearranged in a different format. Here instead of different variables for each time period, the three time periods are stacked into one variable and an additional variable, *Period*, denotes whether the data value is from before, during, or after the pandemic-induced changes to college life.)

## Source

Subigya Nepal, et. al. "Capturing the College Experience: A Four-Year Mobile Sensing Study of Mental Health, Resilience, and Behavior of College Students during the Pandemic", Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 8, 1, Article 38 (March 2024). <https://dl.acm.org/doi/10.1145/3643501>.

---

CricketChirps

*Cricket Chirps*

---

## Description

Cricket chirp rate and temperature

**Format**

A dataset with 7 observations on the following 2 variables.

Temperature	Air temperature in degrees F
Chirps	Cricket chirp rate (chirps per minute)

**Details**

The data were collected by E.A. Bessey and C.A. Bessey who measured chirp rates for crickets and temperatures during the summer of 1898.

**Source**

From E.A Bessey and C.A Bessey, Further Notes on Thermometer Crickets, American Naturalist, (1898) 32, 263-264.

---

 DDS

---

*Developmental Services*


---

**Description**

Funding for individuals by the California Department of Developmental Services (DDS),

**Format**

A dataset with 1000 observations on the following 6 variables.

ID	ID code for subject
AgeCohort	Age group (0-5, 6-12, 13-17, 18-21, 22-50, 50+)
Age	Age in years
Expenditures	Annual expenditures in dollars
Ethnicity	Ethnic group

**Details**

The California Department of Developmental Services (DDS) allocates funds to support developmentally disabled California residents (such as those with autism, cerebral palsy, or intellectual disabilities) and their families. We refer to those supported by DDS as DDS consumers. The dataset DDS includes data on annual expenditure (in \$), ethnicity, age, and gender for 1000 DDS consumers.

**Source**

Taylor, S.A. and Mickel, A. E. (2014). "Simpson's Paradox: A Data Set and Discrimination Case Study Exercise," Journal of Statistics Education, 22(1). The dataset has been altered slightly for privacy reasons, but is based on actual DDS consumers.



---

DecemberFlights	<i>December Flights</i>
-----------------	-------------------------

---

**Description**

Difference between actual and scheduled arrival for United and Delta flights in December 2024.

**Format**

A data frame with 2000 observations on the following 2 variables.

Airline Delta or United

Difference Actual - Scheduled arrival times (in minutes)

**Details**

For a sample of 1000 December flights (in 2024) from each airline, we find the difference between actual and scheduled arrival times. A negative value indicates the flight arrived early.

\*\* Updated for 4e (earlier versions are DecemberFlights3e from 2018 and DecemberFlights2e from 2014.)

**Source**

Downloaded from the Bureau of Transportation Statistics (<https://www.transtats.bts.gov/>).

---

DietDepression	<i>Diet and Depression</i>
----------------	----------------------------

---

**Description**

Results from a study of a short-term diet intervention on depression.

**Format**

A data frame with 75 observations on the following 10 variables.

Group Control or Diet

CESD1 CESD depression score on Day 1

CESD21 CESD depression score on Day 21

CESDDiff Change in CESD depression score

DASS1 DASS depression score on Day 1

DASS21 DASS depression score on Day 21

DASSDiff Change in DASS depression score

BMI1 Body Mass Index on Day 1

BMI21 Body Mass Index on Day 21

BMIDiff Change in Body Mass Index

### Details

A group of researchers in Australia conducted a short (three-week) dietary intervention in a randomized controlled experiment. In the study, 75 college-age students with elevated depression symptoms and relatively poor diet habits were randomly assigned to either a healthy diet intervention group or a control group. The researchers recorded the change over the three-week period on two different numeric scales of depression (the CESD scale and the DASS scale). The CESD (Centre for Epidemiological Studies Depression) score is based more on clinical observations, while the DASS (Depression, Anxiety, and Stress Scale) depends more on self-reported information. They also recorded body mass index (BMI) at the start and end of the 21 day period.

### Source

Francis HM, et al., "A brief diet intervention can reduce symptoms of depression in young adults - A randomised controlled trial," PLoS ONE, 14(10), October 2019.

---

Digits	<i>Digit Counts</i>
--------	---------------------

---

### Description

Digits from social security numbers and student selected "random numbers"

### Format

A dataset with 150 observations on the following 7 variables.

Random	Four digit random numbers given by a sample of students
RND1	First digit
RND2	Second digit
RND3	Third digit
RND4	Fourth digit
SSN8	Eighth digit of social security number
SSN9	Last digit of social security number

### Details

A sample of students were asked to give a random four digit number. The numbers are given in the dataset, along with separate columns for each of the four digits. The data also show the last two digits of each student's social security number (SSN).

### Source

In-class student surveys from several classes.

---

DrugResistance	<i>Drug Resistance</i>
----------------	------------------------

---

**Description**

Effect on drug resistance by level of treatment in mice.

**Format**

A dataset with 72 observations on the following 5 variables.

Treatment	Untreated, Light, Moderate, or Aggressive
Weight	Mouse weight in grams
RBC	Red blood cell density
ResistantDensity	Density of resistant parasites
DaysInfectious	Days infectious with resistant parasites

**Details**

In an experiment to study drug resistance in mice, groups of 18 mice were injected with a mixture of drug-resistant and drug-susceptible malaria parasites. One group received no treatment while the others got limited, moderate, or aggressive amounts of anti-malarial treatment. The weight and red blood cell density reflect the initial health of the mice. Density of resistant parasites and number of days infectious measure the effectiveness of the treatment.

**Source**

Huijben S, Bell AS, Sim DG, Tomasello D, Mideo N, Day T, Read AF (2013) Aggressive chemotherapy and the selection of drug resistant pathogens. PLoS Pathogens 9(9): e1003578.

<http://dx.doi.org/10.1371/journal.ppat.1003578>

Huijben S, et al., (2013). Data from: Aggressive chemotherapy and the selection of drug resistant pathogens. Dryad Digital

Repository. <http://dx.doi.org/10.5061/dryad.09qc0>

---

EducationLiteracy	<i>Education and Literacy</i>
-------------------	-------------------------------

---

**Description**

Education spending and literacy rates for countries.

**Format**

A data frame with 193 observations on the following 4 variables.

Country Name of country

Code Three-letter code for country

Education Education spending (as a percentage of GDP)

Literacy Literacy rate

Details

For each country, we have public spending on education (as a percentage of GDP) and literacy rate (percentage of the adult population who can read and write).  
\*\* Updated for 4e (an earlier versions are at EducationLiteracy2e and EducationLiteracy2e). \*\*

Source

Most recent data (as of 2024) for each country obtained from <https://www.worldbank.org/en/home>.

---

ElectionMargin	<i>Election Margin</i>
----------------	------------------------

---

Description

Approval rating and election margin for recent presidential elections

Format

A dataset with 13 observations on the following 5 variables.

Year	Certain election years from 1940-2020
Candidate	Incumbent US president
Approval	Presidential approval rating at time of election
Margin	Margin of victory/defeat (as a percentage)
Result	Outcome of the election for the incumbent: Lost or Won

Details

Data include US Presidential elections since 1940 in which an incumbent was running for president. The approval rating for the sitting president is compared to the margin of victory/defeat in the election.  
\*\* Updated for 4e with 2020 election, 2e added 2012 election \*\*

Source

Silver, Nate, "Approval Ratings and Re-Election Odds", *fivethirtyeight.com*, posted January 28, 2011 and *http://realclearpolitics.org*

EmployedACS

*Employed in American Community Survey - 2023***Description**

Employed individuals from the American Community Survey (ACS) dataset

**Format**

A data frame with 5862 observations on the following 9 variables.

Sex 0=female and 1=male

Age Age (years)

Married 0=not married and 1=married

Income Wages and salary for the past 12 months (in \$1,000's)

HoursWk Hours of work per week

Race asian, black, other, white

USCitizen 1=citizen and 0=noncitizen

HealthInsurance 1=have health insurance and 0= no health insurance

Language 1=native English speaker and 0=other

**Details**

This is a subset of the ACS dataset from 2023 including only 5862 individuals who were employed. (HoursWk>0)

\*\* Updated for 4e (an earlier versions are EmployedACS2017 and EmployedACS2010). \*\*

**Source**

The full public dataset can be downloaded at <https://www.census.gov/programs-surveys/acs/microdata.html>, and the full list of variables are at <https://www.census.gov/programs-surveys/acs/microdata/documentation.html>

Specific request for these variables is <https://data.census.gov/app/mdat/ACSPUMS1Y2023/table?cv=SEX,MAR,RAC1P,HICOV&vv=AGEP,WAGP,WKHP&rv=CIT,LANX&wt=PWGTP>

ExerciseHours

*Exercise Hours***Description**

Amount of exercise per week for students (and other variables)

**Format**

A data frame with 50 observations on the following 7 variables.

Year Year in school (1=First year,..., 4=Senior)

Sex F or M

Hand Left (l) or Right (r) handed?

Exercise Hours of exercise per week

TV Hours of TV viewing per week

Pulse Resting pulse rate (beats per minute)

Pierces Number of body piercings

**Details**

Data from an in-class survey of statistics students asking about amount of exercise, TV viewing, handedness, sex, pulse rate, and number of body piercings. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.

**Source**

In-class student survey.

---

FacebookFriends

*Facebook Friends*

---

**Description**

Data on number of Facebook friends and grey matter density in brain regions related to social perception and associative memory.

**Format**

A dataset with 40 observations on the following 2 variables.

GMdensity Normalized z-scores of grey matter density in certain brain regions

FBfriends Number of friends on Facebook

**Details**

A recent study in Great Britain examines the relationship between the number of friends an individual has on Facebook and grey matter density in the areas of the brain associated with social perception and associative memory. The study included 40 students at City University London.

**Source**

Kanai, R., Bahrami, B., Roylance, R., and Rees, G., "Online social network size is reflected in human brain structure," *Proceedings of the Royal Society*, 7 April 2012; 279(1732): 1327-1334. Data approximated from information in the article.

FireAnts

*Fire Ants***Description**

Reactions of lizards to the presence of fire ants.

**Format**

A dataset with 80 observations on the following 3 variables.

Invasion	Coded as Uninvaded or Invaded, depending on if the lizard comes from a region with fire ants
Twitches	Number of twitches the lizard makes when encountering fire ants
Flee	Time for the lizard to flee in seconds (more than one minute is recorded as 61).

**Details**

The red imported fire ant, *Solenopsis invicta*, is native to South America, but has an expansive invasive range, including much of the southern United States (invasion of this ant is predicted to go global). In the United States, these ants occupy similar habitats as fence lizards. The ants eat the lizards and the lizards eat the ants, and in either scenario the venom from the fire ant can be fatal to the lizard. The study explored the question of whether lizards learn to adapt their behavior if their environment has been invaded by fire ants by taking lizards from an uninvaded habitat (eastern Arkansas) and lizards from an invaded habitat (southern Alabama, which has been invaded for more than 70 years), exposing them to fire ants, and measuring how long it takes each lizard to flee and the number of twitches each lizard does.

**Source**

Langkilde, T. (2009). "Invasive fire ants alter behavior and morphology of native lizards", *Ecology*, 90(1): 208-217. Thanks to Dr. Langkilde for providing the data.

FisherIris

*Fisher's Iris Data***Description**

Measurements of three iris species

**Format**

A dataset with 150 observations on the following 5 variables.

Type	Species of iris, Setosa, Virginica, or Versicolor
PetalLength	Petal length in mm.
PetalWidth	Petal width in mm.
SepalLength	Sepal length in mm.
SepalWidth	Sepal width in mm.

**Details**

Data used in Fisher's 1936 paper, this famous dataset looks at measurements for samples of three different species of iris. The petal is part of the flower itself and the sepals are green leaves, directly under the petals, providing support.

**Source**

R. A. Fisher (1936). "The use of multiple measurements in taxonomic problems". *Annals of Eugenics* 7 (2): 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x.

---

FishGills12

*Fish Respiration and Calcium - Full Data*

---

**Description**

An experiment to look at fish respiration rates in water with different levels of calcium.

**Format**

A dataset with 360 observations on the following 2 variables.

Calcium	Amount of calcium in the water (mg/L)
GillRate	Respiration rate (beats per minute)

**Details**

Fish were randomly assigned to twelve tanks with different levels (measured in mg/L) of calcium. Respiration rate was measured as number of gill beats per minute.

**Source**

Thanks to Prof. Brad Baldwin for supplying the data.

---

FishGills3

*Fish Respiration and Calcium*

---

**Description**

Respiration rate for fish in three levels of calcium.



**Format**

A dataset with 90 observations on the following 2 variables.

Calcium	Level of calcium Low 0.71 mg/L, Medium 5.24 mg/L, or High 18.24 mg/L
GillRate	Respiration rate (beats per minute)

**Details**

Fish were randomly assigned to three tanks with different levels (low, medium and high) of calcium. Respiration rate was measured as number of gill beats per minute.

**Source**

Thanks to Prof. Brad Baldwin for supplying the data.

---

Flight475	<i>Flight 475</i>
-----------	-------------------

---

**Description**

Flight times for Delta Flight 475 (Boston-SF) in October 2024.

**Format**

A data frame with 31 observations on the following variable.

AirTime Airborne flying time (in minutes) for Flight 475, Boston to San Francisco

**Details**

Delta Airlines Flight 475 was a daily flight from Boston to San Francisco. The data show the airborne flying times for the flight on each day of October 2024.

\*\*Updated for 4e (earlier version are Flight433\_3e (2019) and Flight433\_2e (2016)) \*\*

**Source**

Data collected from the Bureau of Transportation Statistics website at <https://www.transtats.bts.gov/>

---

FloridaLakes

*Florida Lakes*


---

### Description

Water quality measurements for a sample of lakes in Florida

### Format

A dataset with 53 observations on the following 12 variables.

ID	An identifying number for each lake
Lake	Name of the lake
Alkalinity	Concentration of calcium carbonate (in mg/L)
pH	Acidity
Calcium	Amount of calcium in water
Chlorophyll	Amount of chlorophyll in water
AvgMercury	Average mercury level for a sample of fish (large mouth bass) from each lake
NumSamples	Number of fish sampled at each lake
MinMercury	Minimum mercury level in a sampled fish
MaxMercury	Maximum mercury level in a sampled fish
ThreeYrStdMercury	Adjusted mercury level to account for the age of the fish
AgeData	Mean age of fish in each sample

### Details

This dataset describes characteristics of water and fish samples from 53 Florida lakes. Some variables (e.g. Alkalinity, pH, and Calcium) reflect the chemistry of the water samples. Mercury levels were recorded for a sample of large mouth bass selected at each lake.

### Source

Lange, Royals, and Connor, Transactions of the American Fisheries Society (1993)

---

FootballBrain

*Football Brain Measurements*


---

### Description

Brain measurements for non-football players, football players with no concussion history, and football players with a concussion history.

**Format**

A dataset with 75 observations on the following 5 variables.

Group	Control=no football, FBNConcuss=football player but no concussions, or FBConcuss=football player with concussion history
Hipp	Total hippocampus volume, in microL
LeftHipp	Left hippocampus volume, in microL
Years	Number of years playing football
Cognition	Cognitive testing composite reaction time score, given as a percentile

**Details**

The study included 3 groups, with 25 cases in each group. The control group consisted of healthy individuals with no history of brain trauma who were comparable to the other groups in age, sex, and education. The second group consisted of NCAA Division 1 college football players with no history of concussion, while the third group consisted of NCAA Division 1 college football players with a history of concussion. High resolution MRI was used to collect brain hippocampus volume. Data were collected between June 2011 and August 2013. The data values given here are estimated from information given in the paper.

**Source**

Singh R, Meier T, Kuplicki R, Savitz J, et al., "Relationship of Collegiate Football Experience and Concussion With Hippocampal Volume and Cognitive Outcome," JAMA, 311(18), 2014

---

ForestFires

*Forest Fires*


---

**Description**

Characteristics of forest fires in Montesinho park (Portugal)

**Format**

A data frame with 517 observations on the following 13 variables.

X	West to east coordinates for the site (1=farthest west to 9= farthest east)
Y	North to south coordinates for the site (1=farthest north to 9=farthest south)
Month	Month of the year (jan to dec)
Day	Day of the week (sun to sat)
FFMC	Fine fuel moisture code
DMC	Duff moisture code
DC	Drought code
ISI	Initial spread index
Temp	Outside temperature (in celsius)
RH	Relative humidity (in %)
Wind	Wind speed (in km/h)
Rain	Rain in past 30 minutes (in mm/sq-m)
Area	Total burned area (in hectares)

### Details

Data were recorded for fires in the Montesinho natural park in Portugal between January 2000 and December 2003. A map of the park (see the pdf linked below) is divided into 9x9 grid sections (given by the x,y-coordinates in the first two columns of the dataset). There are four components of a Fire Weather Index that rate how weather conditions might increase fire danger. FFMC, DMC, and DC reflect various measures of moisture content, while the ISI score indicated how fast a fire might spread (for example, by wind). For all four measures larger values are associated with more fire danger. Fires that are less than 100 square meters in size (0.01 hectares) are recorded as Area=0.

### Source

Data downloaded from the UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Forest+Fires>

Original article: P. Cortez and A. Morais. "A Data Mining Approach to Predict Forest Fires using Meteorological Data", in New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence (December 2007) <http://www.dsi.uminho.pt/~pcortez/fires.pdf>

---

GeneticDiversity

*Genetic Diversity*

---

### Description

Genetic diversity for different populations are compared to the distance from East Africa.

### Format

A dataset with 52 observations on the following 5 variables.

Population	Identifier for each population
Country	Main country where the population is found
Continent	Continent where the population is found
GeneticDiversity	A measure of genetic diversity in the population
Distance	Distance by land to East Africa (in km)

### Details

The data give a measure of genetic diversity for different populations and the geographic distance of each population from East Africa (Addis Ababa, Ethiopia), as one would travel over the surface of the earth by land (migration long ago is thought to have happened by land).

### Source

Calculated using data from S Ramachandran, O Deshpande, CC Roseman, NA Rosenberg, MW Feldman, LL Cavalli-Sforza. "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa,"" Proceedings of the National Academy of Sciences, 2005, 102: 15942-15947.

---

GlobalInternet2024	<i>Global Internet Usage - 2024</i>
--------------------	-------------------------------------

---

**Description**

Internet usage for several countries in 2024

**Format**

A data frame with 9 observations on the following 3 variables.

Country Name of country

InternetSpeed Average download speed (in Mb)

HoursOnline Average hours online per day for internet users

**Details**

The Worldwide Broadband Speed League tests internet speeds at millions of access points around the world. The average download speed for each country is derived from those data. The DataReportal site provides summaries of country level data on internet usage obtained from various sources. The average number of hours spent online for each country is based on survey data reported at that site.

\*\* Updated for 4e (earlier versions are GlobalInternet2019 and GlobalInternet2011).

**Source**

Internet speeds for 2024 downloaded from <https://www.cable.co.uk/broadband/speed/worldwide-speed-league>

Online hours for 2024 downloaded from <https://datareportal.com/library>

---

GolfRound	<i>Golf Round</i>
-----------	-------------------

---

**Description**

Scorecard for 18 holes of golf

**Format**

A data frame with 18 observations on the following 4 variables.

Hole Hole number (1 to 18)

Distance Length of the hole (in yards)

Par Par for the hole

Score Actual number of strokes needed in this round

**Details**

Data come from a scorecard for one round of golf at the Potsdam Country Club. Par is the expected number of strokes a good golfer should need to complete the hole.

**Source**

Personal file

---

GPABySex

*GPA by Sex*

---

**Description**

Data from a survey of introductory statistics students.

**Format**

A dataset with 343 observations on the following 6 variables.

Exercise	Hours of exercise (per week)
SAT	Combined SAT scores (out of 1600)
GPA	Grade Point Average (0.00-4.00 scale)
Pulse	Pulse rate (beats per minute)
Piercings	Number of body piercings
CodedSex	0=female or 1=male

**Details**

This is a subset of the StudentSurvey dataset where cases with missing values have been dropped and sex is coded as a 0/1 indicator variable.

**Source**

A first day survey over several different introductory statistics classes.

---

GroundhogDay

*Groundhog Day*

---

**Description**

Yearly data on US March temperature and Puxsutawney Phil's forecast on Groundhog Day

**Format**

A data frame with 122 observations on the following 3 variables.

Year Year (1903-2025)

Shadow Phil saw his shadow? (Yes or No)

USMarchTemp Average US temperature in March of that year

Details

Every February 2nd (Groundhog Day) people gather at Gobbler’s Knob in Puxsutawney, Pennsylvania to see if an emerging groundhog (Phil) casts a shadow. Legend has it that a shadow fortells six more weks of winter, but no shadow indicates an early coming of spring. This dataset shows the outcome of Phil’s shadow viewing for each year from 1903 to 2025, along with the average temerature in the US that March. One might expect six more week’s of winter would mean colder temperatures in March.

Source

Data through 2016 from <https://www.kaggle.com/datasets/groundhogclub/groundhog-day> (May 2025).\n\nMore recent years from <https://www.groundhog.org/> and <https://www.ncei.noaa.gov/>.

---

HappyPlanetIndex	<i>Happy Planet Index</i>
------------------	---------------------------

---

Description

Measurements related to happiness and well-being for 147 countries.

Format

A dataset with 147 observations on the following 9 variables.

Country	Name of country
Region	1=Latin America, 2=N. America and Oceania, 3=Western Europe, 4=Middle East and N. Africa, 5=Sub-Saharan Africa, 6=South Asia, 7=Eastern Europe and Central Asia, 8=East Asia
HPI	Happy Planet Index (0-100 scale)
HPIRank	HPI rank for the country
LifeExpectancy	Average life expectancy (in years)
Footprint	Ecological footprint - a measure of the (per capita) ecological impact
LOL	Ladder of Life (a measure of wellbeing on 0-10 scale)
GDPperCapita	Gross Domestic Product (per capita)
Population	Population (in millions)

Details

Data for 147 countries from the Happy Planet Index Project that works to quantify indicators of happiness, well-being, and ecological footprint at a country level.\n\n\*\* Updated for 4e (earlier version is now HappyPlanetIndex2010) \*\*

Source

Abdallah, S. & Hoffman, A. (2024) The Happy Planet Index 2024 Data File.\n\nAccessed from [www.happyplanetindex.org](http://www.happyplanetindex.org) March 2025

---

HeatCognition	<i>Heat and Cognition</i>
---------------	---------------------------

---

### Description

Effect of heat on cognitive ability

### Format

A data frame with 46 observations on the following 3 variables.

AC Whether the student had air conditioning on in the room, No or Yes

MathZRT Z-score of reaction time solving math problems

ColorsZRT Z-score of reaction time solving STROOP color problems

### Details

Forty-six college students were asked to solve cognitive problems first thing in the morning during a heat wave in their Northeastern city. Twenty of the students had air-conditioning in their rooms and twenty-six did not. Z-scores of reaction times are given for math problems and for color dissonance problems.

### Source

Cedeo Laurent JG, Williams A, Oulhote Y, Zanobetti A, Allen JG, Spengler JD "Reduced cognitive function during a heat wave among residents of non-air-conditioned buildings: An observational study of young adults in the summer of 2016." PLoS Med 15(7): e1002605, July 10, 2018. <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002605>. (Dataset is simplified from the repeated measures design used in the original study.)

---

HeightData	<i>Height Data</i>
------------	--------------------

---

### Description

Heights measured for the same 94 children over 18 years.

### Format

A dataset with 94 observations on the following 33 variables.

ID	Identification number)
Sex	M or F
Year_1	Height (in cm.) at age 1 year
Year_1.25	Height (in cm.) at age 1.25 years
Year_1.5	Height (in cm.) at age 1.5 years
Year_1.75	Height (in cm.) at age 1.75 years
Year_2	Height (in cm.) at age 2 years
Year_3	Height (in cm.) at age 3 years



Year_4	Height (in cm.) at age 4 years
Year_5	Height (in cm.) at age 5 years
	See below for full list of years...
Year_17.5	Height (in cm.) at age 17.5 years
Year_18	Height (in cm.) at age 18 years

### Details

In the 1940's and 1950's, the heights of 39 boys and 54 girls, in centimeters, were measured at 30 different time points between the ages of 1 and 18 years as part of the University of California Berkeley growth study. Ages for measurement are 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 7, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18.

### Source

Tuddenham, R. D., and Snyder, M. M. (1954) "Physical growth of California boys and girls from birth to age 18", University of California Publications in Child Development, 1, 183-364.

---

HockeyPenalties2025      *Hockey Penalties (2025)*

---

### Description

Penalty minutes (per game) for NHL teams in 2024-2025

### Format

A data frame with 32 observations on the following 10 variables.

Team Name of the team

W Wins

L Losses

OL Losses in overtime

PTS Points = 2 x W + OL

GF Goals scored

GA Goals allowed

PIM Average penalty minutes per game

OppPIM Average opponent's penalty minutes per game

Playoff Did the team make the playoffs? (N or Y)

### Details

Data give the standings and average number of penalty minutes for each of the 30 National Hockey League (NHL) teams (and their opponents) during the 2024-2025 regular season.

\*\* Updated for 4e (earlier versions are HockeyPenalties2019 and HockeyPenalties2011). \*\*

### Source

Data obtained online at [https://www.hockey-reference.com/leagues/NHL\\_2025.html#all\\_stats](https://www.hockey-reference.com/leagues/NHL_2025.html#all_stats)

HollywoodMovies

*Hollywood Movies - 2019 to 2023***Description**

Data on movies released in Hollywood between 2019 and 2023

**Format**

A data frame with 844 observations on the following 18 variables.

Movie Title of the movie

Distributor Primary U.S. distributor of the movie

Genre Action Adventure, Black Comedy, Comedy, Concert/Performance, Documentary, Drama, Horror, Musical, Romantic Comedy, Thriller/Suspense, Western

RottenTomatoes Rotten Tomatoes rating (critics)

AudienceScore Audience rating (via Rotten Tomatoes)

RunTime Running time (in minutes)

Rating G, PG, PG-13, or R

DomesticBO Box office income for domestic (U.S.) viewers (in millions)

ForeignBO Box office income for foreign viewers (in millions)

WorldwideBO Box office income for all viewers (in millions)

OpenBO Opening weekend box office income (in millions)

TheatersOpen Number of screens for opening weekend

BOAvgOpen Average box office income per theater, opening weekend

Budget Production budget (in millions)

Profitability WorldBO as a percentage of Budget

OpenProfit Percentage of budget recovered on opening weekend

Country Locations of production companies

Year Year the movie was released

**Details**

Information from 844 movies released from Hollywood (or elsewhere in the US) between 2019 and 2023 that had at least \$100,000 in box office income as of February 2025.

\*\* Updated for 4e (earlier versions are HollywoodMovies2018, HollywoodMovies2013, and HollywoodMovies2011). \*\*

**Source**

Movie data obtained in February 2025 from

<https://www.the-numbers.com/>

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

---

HomesForSale	<i>Homes For Sale (2025)</i>
--------------	------------------------------

---

**Description**

Data on homes sold in four states in 2025

**Format**

A data frame with 120 observations on the following 6 variables.

State Location of the home (NY, IA, CA, or FL)

Price Asking price (in \$1,000's)

Size Area of all rooms (in sq. ft.)

Beds Number of bedrooms

Baths Number of bathrooms

Zip Zip code for the home

**Details**

Data for samples of 30 homes recently sold in each state, selected from random zip codes at *zillow.com*.

\*\* Updated for 4e (earlier versions are HomesForSale3e (2019) and HomesForSale2e (2010)). \*\*

**Source**

Data collected from <https://www.zillow.com/> in March 2025.

---

HomesForSaleCA	<i>Homes For Sale in California (2025)</i>
----------------	--

---

**Description**

Data for a sample of homes sold in California

**Format**

A data frame with 30 observations on the following 6 variables.

State Location of the home (CA)

Price Asking price (in \$1,000's)

Size Area of all rooms (in sq. ft.)

Beds Number of bedrooms

Baths Number of bathrooms

Zip Zip code for the home

**Details**

Data for a sample of 30 homes recently sold in California, selected from random zip codes at *zillow.com*. This is a subset of the HomesForSale dataset.

\*\* Updated for 4e (earlier versions are HomesForSaleCA3e (2019) and HomesForSaleCA2e (2010)). \*\*

**Source**

Data collected from <https://www.zillow.com/> in March 2025.

---

HomesForSaleCanton	<i>Homes For Sale in Canton, NY (2025)</i>
--------------------	--

---

**Description**

Data for a sample of homes sold in Canton, NY in 2025

**Format**

A data frame with 10 observations on the following 4 variables.

Price Asking price (in \$1,000's)

Size Area of all rooms (in sq. ft.)

Beds Number of bedrooms

Baths Number of bathrooms

**Details**

Data for a sample of homes recently sold in Canton, NY, selected from *zillow.com*.

\*\* Updated for 4e (earlier versions are HomesForSaleCanton3e (2019) and HomesForSaleCanton2e (2010)). \*\*

**Source**

Data collected from <https://www.zillow.com/> in March 2025.

---

HomesForSaleNY	<i>Homes For Sale in New York (2025)</i>
----------------	--

---

**Description**

Data for a sample of homes sold in New York (state)

**Format**

A data frame with 30 observations on the following 6 variables.

State Location of the home (NY)

Price Asking price (in \$1,000's)

Size Area of all rooms (in sq. ft.)

Beds Number of bedrooms

Baths Number of bathrooms

Zip Zip code for the home

**Details**

Data for a sample of 30 homes recently sold in New York, selected from random zip codes at *zillow.com*. This is a subset of the HomesForSale dataset.

\*\* Updated for 4e (earlier versions are HomesForSaleNY3e (2019) and HomesForSaleNY2e (2010)).  
\*\*

**Source**

Data collected from <https://www.zillow.com/> in 2019.

---

HomingPigeons

*Homing Pigeons*


---

**Description**

Results from the 2019 Midwest Classic Homing Pigeon race

**Format**

A data frame with 1412 observations on the following 5 variables.

Position Finishing position in the race

Loft Name of the pigeon's home loft

Sex C=cock (male) or H=hen (female)

Distance Distance (in miles) from release point to home loft

Speed Speed (in yards per minute)

**Details**

Finishing results from 1412 pigeons completing the 2019 Midwest Classic race for homing pigeons on June 30, 2019. Each loft may enter multiple pigeons.

**Source**

Final race report from the Midwest Homing Pigeon Association, downloaded from <http://www.midwesthpa.com/MIDFinalReports.htm>

Honeybee2023	Honeybee Colonies - 2023
<b>Description</b>	
Number of honeybee colonies (2008-2023)	
<b>Format</b>	
A dataset with 16 observations on the following 2 variables.	
Year	Year
Colonies	Estimated number of honeybee colonies in the US (in thousands)
<b>Details</b>	
Data collected from the USDA on the estimated number of honeybee colonies in the US for the years 2008 through 2023. Updated for 4e (earlier version is now Honeybee2012 with data from 1995-2012).	
<b>Source</b>	
USDA National Agriculture and Statistical Services, <a href="https://quickstats.nass.usda.gov/">https://quickstats.nass.usda.gov/</a> (Accessed February 2025)	
HoneybeeCircuits	Honeybee Circuits

<b>Description</b>	
Number of circuits for honeybee dances and nest quality	
<b>Format</b>	
A dataset with 78 observations on the following 2 variables.	
Circuits	Number of waggle dance circuits for a returning scout bee
Quality	Quality of the nest site: High or Low
<b>Details</b>	
When honeybees are looking for a new home, they send out scouts to explore options. When a scout returns, she does a "waggle dance" with multiple circuit repetitions to tell the swarm about the option she found. The bees then decide between the options and pick the best one. Scientists wanted to find out how honeybees decide which is the best option, so they took a swarm of honeybees to an island with only two possible options for new homes: one of very high honeybee quality and one of low quality. They then kept track of the scouts who visited each option and counted the number of waggle dance circuits each scout bee did when describing the option.	

**Source**

Seeley, T., Honeybee Democracy, Princeton University Press, Princeton, NJ, 2010, p. 128

---

HoneybeeWaggle

*Honeybee Waggle*

---

**Description**

Honeybee dance duration and distance to nesting site

**Format**

A dataset with 7 observations on the following 2 variables.

Distance	Distance to the potential nest site (in meters)
Duration	Duration of the waggle dance (in seconds)

**Details**

When honeybee scouts find a food source or a nice site for a new home, they communicate the location to the rest of the swarm by doing a "waggle dance." They point in the direction of the site and dance longer for sites farther away. The rest of the bees use the duration of the dance to predict distance to the site.

**Source**

Seeley, T., Honeybee Democracy, Princeton University Press, Princeton, NJ, 2010, p. 128

---

HotDogs2024

*Hot Dog Eating Contest - 2024*

---

**Description**

Winning number of hot dogs consumed in an eating contest (2002-2024)

**Format**

A data frame with 23 observations on the following 2 variables.

Year Year of the contest: 2002 to 2024

HotDogs Winning number of hot dogs consumed

**Details**

Every Fourth of July, Nathan's Famous in New York City holds a hot dog eating contest, in which contestants try to eat as many hot dogs (with buns) as possible in ten minutes. The winning number of hot dogs are given for each year from 2002-2024.

\*\* Data set updated for 4e (earlier versions are HotDogs2019, HotDogs2015, and HotDogs1e) \*\*

**Source**

Downloaded from [https://en.wikipedia.org/wiki/Nathan's\\_Hot\\_Dog\\_Eating\\_Contest](https://en.wikipedia.org/wiki/Nathan's_Hot_Dog_Eating_Contest)

---

HouseStarts2024	<i>Housing Starts (2000-2024)</i>
-----------------	-----------------------------------

---

### Description

Quarterly housing starts in the United States from 2000-2024

### Format

A data frame with 100 observations on the following 3 variables.

Year Year (2000 to 2024)

Quarter Q1=Jan-Mar, Q2=Apr-June, Q3=July-Sept, Q4=Oct-Dec

Houses New US residential house construction starts (in thousands)

### Details

Number of new homes started in the US for each quarter from 2000-2024.

Updated for 4e (earlier version are in HouseStarts2018 and HouseStarts2015)

### Source

Census.gov website <https://www.census.gov/econ/currentdata/>

<https://www.census.gov/econ/currentdata/dbsearch?program=RESCONST&startYear=2000&endYear=2024&categories=STARTS&dataType=SINGLE&geoLevel=US&notAdjusted=1&submit=GET+DATA&releaseScheduleId=>

---

HumanTears25	<i>Human Tears -Sadness and Sexual Arousal</i>
--------------	--

---

### Description

Differences in sadness and sexual arousal ratings for 25 men sniffing female tears or a placebo in a matched pairs experiment.

### Format

A data frame with 25 observations on the following 2 variables.

SexDiff Difference in sexual arousal rating (placebo rating - tears rating)

SadDiff Difference in sadness rating (placebo rating - tears rating)



**Details**

Twenty-five men had a pad attached to their upper lip that contained either female tears collected from women who watched a sad film or a salt solution (as a placebo) that had been trickled down the same women's faces. The data were collected following a double-blind matched pairs design, where the order was randomized. The men were shown pictures of female faces and asked "To what extent is this face sad?" or "To what extent is this face sexually arousing?" Men's answers were input using a Visual Analog Scale, which were then converted to a scale with results between about 200 and 800. The data show the difference in rating (placebo rating minus sadness rating) for each man for the sad question (SadDiff) or the sexual arousal question (SexDiff). Data are approximated from information given in the article.

**Source**

Gelstein, S, et al., "Human Tears Contain a Chemosignal," Science, 331(6014), 226-230, January 14, 2011.

---

HumanTears50

*Human Tears - Testosterone*

---

**Description**

Differences in testosterone levels for 50 men in a matched pairs experiment, where the differences are between sniffing female tears and sniffing a placebo

**Format**

A data frame with 50 observations on the following 3 variables.

Placebo Testosterone level after sniffing a placebo

Tears Testosterone level after sniffing female tears

Difference Difference in testosterone level (Placebo - Tears)

**Details**

Fifty men had a pad attached to their upper lip that contained either female tears collected from women who watched a sad film or a salt solution (as a placebo) that had been trickled down the same women's faces. The data were collected following a double-blind matched pairs design, where the order was randomized and the data were collected on consecutive days. After sniffing each substance (placebo or tears), men had their salivary testosterone levels measured, in pg/ml. Data are approximated from information given in the article.

**Source**

Gelstein, S, et al., "Human Tears Contain a Chemosignal," Science, 331(6014), 226-230, January 14, 2011.

---

Hurricanes2024	<i>Hurricanes (1914 to 2024)</i>
----------------	----------------------------------

---

**Description**

Hurricanes in the North Atlantic each year (1914-2024)

**Format**

A data frame with 111 observations on the following 2 variables.

Year Year (1914 to 2024)

Hurricanes Number of North Atlantic hurricanes

**Details**

Number of North Atlantic hurricanes - yearly 1914-2024.

\*\* Updated for 4e (earlier versions are in Hurricanes2018 and Hurricanes2014). \*\*

**Source**

Weather Underground website at <https://www.wunderground.com/hurricane/archive>

---

ICUAdmissions	<i>Intensive Care Unit Admissions</i>
---------------	---------------------------------------

---

**Description**

Data from patients admitted to an intensive care unit

**Format**

A dataset with 200 observations on the following 21 variables.

ID	Patient ID number
Status	Patient status: 0=lived or 1=died
Age	Patient's age (in years)
Sex	0=male or 1=female
Race	Patient's race: 1=white, 2=black, or 3=other
Service	Type of service: 0=medical or 1=surgical
Cancer	Is cancer involved? 0=no or 1=yes
Renal	Is chronic renal failure involved? 0=no or 1=yes
Infection	Is infection involved? 0=no or 1=yes
CPR	Patient gets CPR prior to admission? 0=no or 1=yes
Systolic	Systolic blood pressure (in mm of Hg)
HeartRate	Pulse rate (beats per minute)
Previous	Previous admission to ICU within 6 months? 0=no or 1=yes
Type	Admission type: 0=elective or 1=emergency
Fracture	Fractured bone involved? 0=no or 1=yes

P02	Partial oxygen level from blood gases under 60? 0=no or 1=yes
PH	pH from blood gas under 7.25? 0=no or 1=yes
PC02	Partial carbon dioxide level from blood gas over 45? 0=no or 1=yes
Bicarbonate	Bicarbonate from blood gas under 18? 0=no or 1=yes
Creatinine	Creatinine from blood gas over 2.0? 0=no or 1=yes
Consciousness	Level: 0=conscious, 1=deep stupor, or 2=coma

### Details

Data from a sample of 200 patients following admission to an adult intensive care unit (ICU).

### Source

DASL dataset downloaded from <http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html>

---

ImmuneTea	<i>Immune Tea</i>
-----------	-------------------

---

### Description

Interferon gamma production and tea drinking

### Format

A dataset with 21 observations on the following 2 variables.

InterferonGamma	Measure of interferon gamma production
Drink	Type of drink: Coffee or Tea

### Details

Eleven healthy non-tea-drinking individuals were asked to drink five or six cups of tea a day, while ten healthy non-tea and non-coffee-drinkers were asked to drink the same amount of coffee, which has caffeine but not the L-theanine that is in tea. The groups were randomly assigned. After two weeks, blood samples were exposed to an antigen and production of interferon gamma was measured.

### Source

Adapted from Kamath, et.al., "Antigens in tea-Beverage prime human V 2V2 T cells in vitro and in vivo for memory and non-memory antibacterial cytokine responses", Proceedings of the National Academy of Sciences, May 13, 2003.

---

InkjetPrinters	<i>Inkjet Printers</i>
----------------	------------------------

---

### Description

Data from online reviews of inkjet printers (2025)

### Format

A dataset with 25 observations on the following 7 variables.

Model	Model name of printer
PPM	Printing rate (pages per minute) for a benchmark set of print jobs
PhotoTime	Time (in seconds) to print 4x6 color photos
Price	Typical retail price (in dollars)
CostBW	Cost per page (in cents) for printing in black & white
CostColor	Cost per page (in cents) for printing in color
Year	Year printer was first released

### Details

Information from reviews of all-one inkjet printers at Rtings.com in January 2025.

### Source

Inkjet printer data found at <https://www.rtings.com/printer/tools/table/157814>, January 2025.

---

LifeExpectancyVehicles	<i>Life Expectancy and Vehicle Registrations (2022)</i>
------------------------	---

---

### Description

Yearly US life expectancy and number of registered vehicles (1970-2022)

### Format

A data frame with 53 observations on the following 3 variables.

Year	Year (1970 to 2022)
LifeExpectancy	Average life expectancy (in years) for babies born in the year
Vehicles	Number of motor vehicles registered in the US (in millions)

### Details

Life expectancy (in years for babies born each year) and number of vehicles registered in the US for each year from 1970 to 2022.

\*\* Updated for 4e (earlier versions are LifeExpectancyVehicles2e, LifeExpectancyVehicles2e, and LifeExpectancyVehicles1e) \*\*

**Source**

Vehicle registrations from the Federal Highway Administration, <https://www.fhwa.dot.gov/policyinformation/statistics.cfm>.

Lifetime data from the Centers for Disease Control and Prevention, National Center for Health Statistics <https://www.cdc.gov/nchs/fastats/life-expectancy.htm>.

---

 LightatNight

*Light at Night for Mice*


---

**Description**

Data on body mass gain from an experiment with mice having different nighttime light conditions

**Format**

A dataset with 18 observations on the following 2 variables.

Group	Light=dim light at night or Dark=dark at night
BMGain	Body mass gain (in grams over a three week period)

**Details**

In this study, 18 mice were randomly split into two groups. One group was on a normal light/dark cycle (Dark) and the other group had light during the day and dim light at night (Light). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice with dim light at night, however, consumed much of their food during the well-lit rest period, when most mice are usually sleeping. The change in body mass was recorded after three weeks.

\*\* See also LightatNight4Weeks or LightatNight8Weeks for more variables measured at other points in the same experiment, with a third experimental condition which had 9 additional mice with a bright light on all the time. \*\*

**Source**

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

LightatNight4Weeks	<i>Light at Night for Mice - After 4 Weeks</i>
--------------------	--

---

### Description

Data from an experiment with mice having different nighttime light conditions

### Format

A dataset with 27 observations on the following 9 variables.

Light	DM=dim light at night, LD=dark at night, or LL=bright light at night
BMGain	Body mass gain (in grams over a four week period)
Corticosterone	Blood corticosterone level (a measure of stress)
DayPct	Percent of calories eaten during the day
Consumption	Daily food consumption (grams)
GlucoseInt	Glucose intolerant? No or Yes
GTT15	Glucose level in the blood 15 minutes after a glucose injection
GTT120	Glucose level in the blood 120 minutes after a glucose injection
Activity	A measure of physical activity level

### Details

In this study, 27 mice were randomly split into three groups. One group was on a normal light/dark cycle (LD), one group had bright light on all the time (LL), and one group had light during the day and dim light at night (DM). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice in both dim light and bright light, however, consumed more than half of their food during the well-lit rest period, when most mice are sleeping. Values in this dataset are recorded after four weeks in the experimental condition.

\*\* This dataset was named LightatNight in the first edition \*\*

\*\* See also LightatNight8Weeks for the same data after 8 weeks or LightatNight with just BMGain after 3 weeks for the DM and LD groups. \*\*

### Source

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

LightatNight8Weeks	<i>Light at Night for Mice - After 8 Weeks</i>
--------------------	--

---

### Description

Data from an experiment with mice having different nighttime light conditions

**Format**

A dataset with 27 observations on the following 9 variables.

Light	DM=dim light at night, LD=dark at night, or LL=bright light at night
BMGain	Body mass gain (in grams over an eight week period)
Corticosterone	Blood corticosterone level (a measure of stress)
DayPct	Percent of calories eaten during the day
Consumption	Daily food consumption (grams)
GlucoseInt	Glucose intolerant? No or Yes
GTT15	Glucose level in the blood 15 minutes after a glucose injection
GTT120	Glucose level in the blood 120 minutes after a glucose injection
Activity	A measure of physical activity level

**Details**

In this study, 27 mice were randomly split into three groups. One group was on a normal light/dark cycle (LD), one group had bright light on all the time (LL), and one group had light during the day and dim light at night (DM). The dim light was equivalent to having a television set on in a room. The mice in darkness ate most of their food during their active (nighttime) period, matching the behavior of mice in the wild. The mice in both dim light and bright light, however, consumed more than half of their food during the well-lit rest period, when most mice are sleeping. Values in this dataset are recorded after eight weeks in the experimental condition.

\*\* See also `LightatNight4Weeks` for the same data after 4 weeks or `LightatNight` with just `BMGain` after 3 weeks for just the DM and LD groups. \*\*

**Source**

Fonken, L., et. al., "Light at night increases body mass by shifting time of food intake," Proceedings of the National Academy of Sciences, October 26, 2010; 107(43): 18664-18669.

---

MalevolentUniformsNFL    *Malevolent Uniforms NFL*

---

**Description**

Perceived malevolence of uniforms and penalties for National Football League (NFL) teams

**Format**

A dataset with 28 observations on the following 3 variables.

NFLTeam	Team name
NFL_Malevolence	Score reflecting the "malevolence" of a team's uniform
ZPenYds	Z-score for penalty yards

**Details**

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty yards converted to z-scores and averaged for each team over the seasons from 1970-1986.

**Source**

Frank and Gilovich, "The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports", Journal of Personality and Social Psychology, Vol. 54, No. 1, 1988, p. 74-85.

---

MalevolentUniformsNHL    *Malevolent Uniforms NHL*

---

**Description**

Perceived malevolence of uniforms and penalties for National Hockey League (NHL) teams

**Format**

A dataset with 28 observations on the following 3 variables.

NHLTeam	Team name
NHL_Malevolence	Score reflecting the "malevolence" of a team's uniform
ZPenMin	Z-score for penalty minutes

**Details**

Participants with no knowledge of the teams rated the jerseys on characteristics such as timid/aggressive, nice/mean and good/bad. The averages of these responses produced a "malevolence" index with higher scores signifying impressions of more malevolent uniforms. To measure aggressiveness, the authors used the amount of penalty minutes converted to z-scores and averaged for each team over the seasons from 1970-1986.

**Source**

Frank and Gilovich, "The Dark Side of Self- and Social Perception: Black Uniforms and Aggression in Professional Sports", Journal of Personality and Social Psychology, Vol. 54, No. 1, 1988, p. 74-85.

---

MammalLongevity    *Mammal Longevity*

---

**Description**

Longevity and gestation period for mammals



**Format**

A dataset with 40 observations on the following 3 variables.

Animal	Species of mammal
Gestation	Time from fertilization until birth (in days)
Longevity	Average lifespan (in years)

**Details**

Dataset with average lifespan (in years) and typical gestation period (in days) for 40 different species of mammals.

**Source**

2010 World Almanac, pg. 292.

---

ManhattanApartments	<i>Manhattan Apartment Prices (2025)</i>
---------------------	--

---

**Description**

Apartment prices for sale in Manhattan in 2025

**Format**

A data frame with 20 observations on the following variable.

Rent Monthly rent (in dollars)

**Details**

Monthly rents for a sample of 20 one-bedroom apartments in Manhattan, NY in March 2025

**Source**

Apartments newly available on Zillow at <https://www.zillow.com/manhattan-new-york-ny/rentals/>, March, 2025.

---

MarriageAges

*Marriage Ages*

---

**Description**

Ages for husbands and wives from marriage licenses

**Format**

A dataset with 100 observations on the following 2 variables.

Husband	Age of husband at marriage
Wife	Age of wife at marriage

**Details**

Data from a sample of 100 marriage licenses in St. Lawrence County, NY gives the ages of husbands and wives for newly married couples.

**Source**

Thanks to Linda Casserly, St. Lawrence County Clerk's Office

---

MastersGolf

*Masters Golf Scores*

---

**Description**

Scores from the 2011 Masters golf tournament

**Format**

A dataset with 20 observations on the following 2 variables.

First	First round score (in relation to par)
Final	Final four round score (in relation to par)

**Details**

Data for a random sample of 20 golfers who made the cut at the 2011 Masters golf tournament.

**Source**

2011 Masters tournament results at [http://www.masters.com/en\\_US/discover/past\\_winners.html](http://www.masters.com/en_US/discover/past_winners.html)

---

MateChoice

*Fruitfly Survival - by Mate Choice*


---

**Description**

Number of fruitflies surviving depending on number of mating choices.

**Format**

A dataset with 50 observations on the following 3 variables.

Choice	Number of surviving larvae (out of 200) when female had a choice of mates
NoChoice	Number of surviving larvae (out of 200) when female had only one choice for a mate
Difference	Choice - NoChoice

**Details**

In an experiment, two hundred larvae from female fruitflies that were exposed to many male fruitflies were tracked to see how many survived. This was compared to a different set of 200 larvae from females that were exposed to only one male each. Values in the dataset give how many of the 200 larvae survived. This process was replicated 50 times, so each row of the dataset corresponds to the survival counts (and difference) for one run, starting with 200 larvae of each type.

**Source**

Patridge, L. (1980). "Mate choice increases a component of offspring fitness in fruit flies," *Nature*, 283:290-291, 1/17/80.

---

MentalMuscle

*Mental Muscle*


---

**Description**

Comparing actual movements to mental imaging movements

**Format**

A dataset with 32 observations on the following 3 variables.

Action	Treatment: Actual motions or Mental imaging motions
PreFatigue	Time (in seconds) to complete motions before fatigue
PostFatigue	Time (in seconds) to complete motions after fatigue

**Details**

In this study, participants were asked to either perform actual arm pointing motions or to mentally imagine equivalent arm pointing motions. Participants then developed muscle fatigue by holding a heavy weight out horizontally as long as they could. After becoming fatigued, they were asked to repeat the previous mental or actual motions. Eight participants were assigned to each group, and the time in seconds to complete the motions was measured before and after fatigue.

**Source**

Data approximated from summary statistics in: Demougeot L. and Papaxanthis C., "Muscle Fatigue Affects Mental Simulation of Action," The Journal of Neuroscience, July 20, 2011, 31(29):10712-10720.

---

MindsetMatters

*Mindset Matters*


---

**Description**

Data from a study of perceived exercise with maids

**Format**

A dataset with 75 observations on the following 14 variables.

Cond	Treatment condition: 0=uninformed or 1=informed
Age	Age (in years)
Wt	Original weight (in pounds)
Wt2	Weight after 4 weeks (in pounds)
BMI	Original body mass index
BMI2	Body mass index after 4 weeks
Fat	Original body fat percentage
Fat2	Body fat percentage after 4 weeks
WHR	Original waist to hip ratio
WHR2	Waist to hip ratio after 4 weeks
Syst	Original systolic blood pressure
Syst2	Systolic blood pressure after 4 weeks
Diast	Original diastolic blood pressure
Diast2	Diastolic blood pressure after 4 weeks

**Details**

In 2007 a Harvard psychologist recruited 75 female maids working in different hotels to participate in a study. She informed 41 maids (randomly chosen) that the work they do satisfies the Surgeon General's recommendations for an active lifestyle (which is true), giving them examples for how and why their work is good exercise. The other 34 maids were told nothing (uninformed). Various characteristics (weight, body mass index, ...) were recorded for each subject at the start of the experiment and again four weeks later. Maids with missing values for weight change have been removed.

**Source**

Crum, A.J. and Langer, E.J. (2007). Mind-Set Matters: Exercise and the Placebo Effect, Psychological Science, 18:165-171. Thanks to the authors for supplying the data.

MustangPrice

*Mustang Prices***Description**

Price, age, and mileage for used Mustang cars at an internet website (2025)

**Format**

A dataset with 30 observations on the following 3 variables.

Age	Age of the car (in years)
Miles	Mileage on the car (in 1,000's)
Price	Asking price (in \$1,000's)

**Details**

A statistics student was interested in prices for used Mustang cars being offered for sale on an internet site. He sampled 30 cars from the website and recorded the age (in years), mileage (in thousands of miles), and asking price (in \$1,000's) for each car in his sample.

\*\* Updated for 4e (earlier version is now MustangPrice1e). \*\*

**Source**

Student project with data collected from *autotrader.com* in March, 2025.

NBAPlayers2024

*NBA Players Data for 2023-2024 Season***Description**

Data from the 2023-2024 regular season for 237 NBA basketball players.

**Format**

A data frame with 237 observations on the following 27 variables.

Player Name of player

Pos PG=point guard, SG=shooting guard, PF=power forward, SF=small forward, C=center

Age Age (in years)

Team Team name (2TM or 3TM = players that switched teams during the season)

Games Games played (out of 82)

Starts Games started

Mins Minutes played

MinPerGame Minutes per game

FGMade Field goals made

FGAttempt Field goals attempted  
 FGPct Field goal percentage  
 FG3Made Three-point field goals made  
 FG3Attempt Three-point field goals attempted  
 FG3Pct Three-point field goal percentage  
 FTMade Free throws made  
 FTAttempt Free throws attempted  
 FTPct Free throw percentage  
 OffRebounds Offensive rebounds  
 DefRebounds Defensive rebounds  
 Rebounds Total rebounds  
 Assists Number of assists  
 Steals Number of steals  
 Blocks Number of blocked shots  
 Turnovers Number of turnovers  
 Fouls Number of personal fouls  
 Points Number of points scored  
 PPG Points per game played

### Details

Data for 237 NBA basketball players from the 2023-2024 regular season. Includes all players who played in at least 20 games and averaged at least 20 minutes per game that season.

\*\* Data set updated for 4e (earlier versions are NBAPlayers2019, NBAPlayers2015, and NBAPlayers2011). \*\*

### Source

[https://www.basketball-reference.com/leagues/NBA\\_2024\\_totals.html](https://www.basketball-reference.com/leagues/NBA_2024_totals.html)

---

NBAStandings2024

*NBA 2023-2024 Regular Season Standings*

---

### Description

Won-Loss record and statistics for NBA Teams in 2023-2024

### Format

A data frame with 30 observations on the following 6 variables.

Team Team name  
 Wins Number of wins in an 82 game regular season  
 Losses Number of losses  
 WinPct Proportion of games won  
 PtsFor Average points scored per game  
 PtsAgainst Average points allowed per game

**Details**

Won-Loss record and regular season statistics for 30 teams in the National Basketball Association for the 2023-2024 season.

\*\* Data set updated for 4e (earlier version are NBAStandings2019, NBAStandings2016, and NBAStandings1e) \*\*

**Source**

Data downloaded from [http://www.basketball-reference.com/leagues/NBA\\_2024.html](http://www.basketball-reference.com/leagues/NBA_2024.html)

---

NFLContracts2024	<i>NFL Contracts in 2024</i>
------------------	------------------------------

---

**Description**

Dollar size of contracts for all NFL players in 2024

**Format**

A data frame with 2505 observations on the following 5 variables.

Player Player's name

Position Code for the primary position of the player (QB=quarterback, etc.)

Team Nickname of the team

TotalMoney Total value of the contract (in millions of dollars)

YearlySalary Average salary per year (in millions of dollars)

**Details**

This dataset contains salary information for all National Football League (NFL) players under contract for the 2024 season. Many contracts extend over multiple years, so TotalMoney gives the overall size of the contract and YearlySalary indicates the average yearly salary. All amounts are in millions of dollars.

\*\* Updated for 4e (earlier versions are NFLContracts2019 and NFLContracts2015). \*\*

**Source**

Contract data collected from <https://overthecap.com>, accessed February, 2025.

NFLPreseason2024

*Wins for NFL Teams (2021-2024)***Description**

Number of preseason and regular season wins for NFL teams, each year from 2021 to 2024.

**Format**

A data frame with 128 observations on the following 10 variables.

Season Year between 2021 and 2024

Team Name of one of 32 NFL teams

PreWins Number of preseason wins (out of 3 games)

PrePF Points scored in the preseason

PrePA Points allowed in the preseason

PreDiff Points scored minus points allowed in the preseason

RegWins Number of regular season wins (out of 17 games)

RegPF Points scored in the regular season

RegPA Points allowed in the regular season

RegDiff Points scored minus points allowed in the regular season

**Details**

Number of wins, points scored and allowed in the preseason (out of 3 preseason games) and regular season (out of 17 regular season games) for each of the 32 National Football (NFL) teams over a four year period from 2021 to 2024. Ties count as one half win.

\*\* Updated for 4e (earlier versions or seasons with with 4 preseason and 16 regular season games are now NFLPreseason2019 and NFLPreseason2014). \*\*

**Source**

Data available at <https://www.pro-football-reference.com/>.

NFLScores2024

*NFL Scores in 2024***Description**

Results for all NFL games for the 2024 regular season



**Format**

A data frame with 272 observations on the following 11 variables.

Week Week of the season (1 through 18)  
 HomeTeam Home team name  
 AwayTeam Visiting team name  
 HomeScore Points scored by the home team  
 AwayScore Points scored by the visiting team  
 HomeYards Yards gained by the home team  
 AwayYards Yards gained by the visiting team  
 HomeTO Turnovers lost by the home team  
 AwayTO Turnovers lost by the visiting team  
 Date Date of the game  
 Day Day of the week (Mon, Sat, Sun, or Thu)

**Details**

Data for all 272 regular season games in the National Football League (NFL) for the 2024 season.  
 \*\* Updated for 4e (earlier versions are NFLScores2018 and NFLScores2011). \*\*

**Source**

NFL scores and game statistics found at <https://www.pro-football-reference.com/years/2024/games.htm>.

---

 NHANES

---

*National Health and Nutrition Examination Survey (NHANES) Subset*


---

**Description**

A subset of the 2009-2010 National Health and Nutrition Examination Survey (NHANES).

**Format**

A data frame with 4716 observations on the following 5 variables.

Case Case ID number  
 Organic Buy any food labeled organic (past 30 days)? (No or Yes)  
 Health Self-rating of health (Excellent, Very good, Fair, Good, or Poor)  
 HealthBinary Health with two categories: Poor / Fair / Good or Very good / Excellent  
 Income Monthly income? (dollars)

**Details**

This dataset is a subset of the 2009-2010 National Health and Nutrition Examination Survey (NHANES). NHANES is a national survey conducted by the Centers for Disease Control and Prevention (CDC) on a random sample of Americans. This subset contains data on select variables for the subset of people with responses to the questions about buying organic food and self-reported health status.

### Source

The data were downloaded from <https://www.cdc.gov/nchs/nhanes/index.htm>.

---

Ninja	<i>Ninja Warrior Times</i>
-------	----------------------------

---

### Description

Times for students doing a Ninja Warrior obstacle course

### Format

A data frame with 99 observations on the following 9 variables.

Name First name of the child

Age Age of the child (in years)

Sex Sex of the child (F or M)

Time1 Time to complete the course, in seconds, the first time the child was timed on the course

Time2 Time, in seconds, for a second run through the course

Time3 Time, in seconds, for a third run through the course

Time4 Time, in seconds, for a fourth run through the course

Time5 Time, in seconds, for a fifth run through the course

AvgTime Average time for all runs the child completed

### Details

"Ninja Warrior" is a new sport based off the TV show "American Ninja Warrior" in which people go through an obstacle course as fast as they can. The sport is growing in popularity and set to debut at the 2028 Summer Olympic Games under the name "Obstacle Course Racing" as part of the Modern Pentathlon. \

This dataset contains times from a 2025 Ninja Warrior class for children. Due to time constraints some children only did the course once, while others were timed for up to five different runs.

### Source

Data collected in 2025 from a class at Centre Ninja, <https://www.centreeelitegymnastics.com/centre-ninja>. Thanks to Coach Brett Corbishley for collecting the data, and to Axel and Cal Lock Morgan for entering the data.

---

NutritionStudy	<i>Nutrition Study</i>
----------------	------------------------

---

## Description

Variables related to nutrition and health for 315 individuals

## Format

A dataset with 315 observations on the following 17 variables.

ID	ID number for each subject in this sample
Age	Subject's age (in years)
Smoke	Smoker? coded as No or Yes
Quetelet	Weight/(Height <sup>2</sup> )
Vitamin	Vitamin use: coded as 1=Regularly, 2=Occasionally, or 3=No
Calories	Number of calories consumed per day
Fat	Grams of fat consumed per day
Fiber	Grams of fiber consumed per day
Alcohol	Number of alcoholic drinks consumed per week
Cholesterol	Cholesterol consumed (mg per day)
BetaDiet	Dietary beta-carotene consumed (mcg per day)
RetinolDiet	Dietary retinol consumed (mcg per day)
BetaPlasma	Plasma beta-carotene (ng/ml)
RetinolPlasma	Plasma retinol (ng/ml)
Sex	Coded as Female or Male
VitaminUse	Coded as No Occasional Regular
PriorSmoke	Smoking status: coded as 1=Never, 2=Former, or 3=Current

## Details

Data from a cross-sectional study to investigate the relationship between personal characteristics and dietary factors, and plasma concentrations of retinol, beta-carotene and other carotenoids. Study subjects were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous.

## Source

Nierenberg, Stukel, Baron, Dain, and Greenberg, "Determinants of plasma levels of beta-carotene and retinol", American Journal of Epidemiology (1989). Data downloaded from [http://lib.stat.cmu.edu/datasets/Plasma\\_Retinol](http://lib.stat.cmu.edu/datasets/Plasma_Retinol).

---

OlympicMarathon2024	<i>2024 Olympic Men's Marathon</i>
---------------------	------------------------------------

---

### Description

Times for all finishers in the men's marathon at the 2024 Olympics

### Format

A data frame with 71 observations on the following 4 variables.

Athlete Name of marathoner

Country Nationality of marathoner (3 letter country code)

Time Time as H:MM:SS

Minutes Time in minutes

### Details

Results for all finishers in the 2024 Men's Olympic marathon in Paris, France.

\*\* Updated for 4e (earlier versions are now in OlympicMarathon 2016, OlympicMarathon2012, and OlympicMarathon2008) \*\*

### Source

<https://www.olympics.com/en/olympic-games/paris-2024/results/athletics/men-marathon>

---

OrganicEffect	<i>Eating Organic Foods</i>
---------------	-----------------------------

---

### Description

Data comparing pesticide levels in family members when eating non-organic vs organic food

### Format

A dataset with 160 observations on the following 6 variables.

Person	Code for family member, Father, Mother, GirlA, GirlB, Boy
Pesticide	One of eight different pesticides measured
Day	Day of the measurement (Day1, Day3, Day4, or Day6)
NonOrganic	Level of the pesticide after eating a non-organic diet
Organic	Level of the pesticide after eating an organic diet
Diff	Difference = NonOrganic - Organic

## Details

A study looked at a Swedish family that ate a conventional diet (non-organic), and then had them eat only organic for two weeks. Pesticide concentrations for several different pesticides were measured in micrograms/g creatinine by testing morning urine. Multiple measurements were taken for each person before the switch to organic foods, and then again after participants had been eating organic for at least one week.

## Source

Magner, J., Wallberg, P., Sandberg, J., and Cousins, A.P. (2015). "Human exposure to pesticides from food: A pilot study," IVL Swedish Environmental Research Institute.  
[https://www.coop.se/PageFiles/429812/Coop%20Ekoeffekten\\_Report%20ENG.pdf](https://www.coop.se/PageFiles/429812/Coop%20Ekoeffekten_Report%20ENG.pdf), January 2015

---

OttawaSenators2024

*Ottawa Senators Hockey Team (2023-2024)*

---

## Description

Data for 23 players on the 2023-2024 Ottawa Senators NHL team

## Format

A data frame with 23 observations on the following 10 variables.

Player Players name

Position D=defense, C=center, RW=right wing, LW=left wing

Age Age (in years)

Games Games played in the 2023-24 NHL season (out of 82)

Goals Goals

Assists Assists

Points Goals + Assists

PlusMinus Difference between (even strength) goals for and against while on ice

PenMins Number of penalty minutes

MinPerGame Average minutes on the ice per game

## Details

Data for all players (except goalies) who played at least 20 games with the Ottawa Senators hockey team in the 2023-2024 NHL season.

\*\* Updated for 4e (previous versions are now OttawaSenators2019, OttawaSenators2015, and OttawaSenators1e) \*\*

## Source

Data downloaded from <https://www.hockey-reference.com/teams/OTT/2024.html> (March 2025)

PASeniors

*Pennsylvania High School Seniors (2023-2024)***Description**

Information on a sample of high school seniors from the state of Pennsylvania between 2023 and 2024.

**Format**

A data frame with 500 observations on the following 35 variables.

Year Year student submitted data (2023 or 2024)

Gender Female or Male

Age Age (in years)

Hand Dominant hand (Left, Right, or Both)

Height Height (in cm)

Foot Foot length (in cm)

Armspan Armspan (in cm)

Languages Languages spoken

GetToSchool Main mode of transportation to school (Bus, Car, or Walk - Walk includes bicycle)

TravelTime Travel time to school (in minutes)

ReactionTime Time (in seconds) to click when a color changes

MemoryScore Score in an online memory game

Activity Favorite physical activity

Music Favorite genre of music

BirthMonth Birth month

Season Favorite season

Allergies Have allergies? (No or Yes)

Vegetarian Vegetarian? (No or Yes)

FavFood Favorite food

FavDrink Beverage used most often during the day

FavSubject Favorite subject in school

Sleep1 Typical hours of sleep on a school night

Sleep2 Typical hours of sleep on a non-school night

Occupants Number of occupants at home

Communicate Most often method to communicate with friends

TextsSent Number of texts sent (previous day)

HangHours Hours last week spent hanging out with friends

HWHours Hours last week spent doing homework

VideoGameHours Hours last week spent playing computer/video games

ComputerHours Hours last week spent using a computer

- TVHours Hours last week spent watching TV
- WorkHours Hours last week spent working at a paid job
- SchoolPressure Amount of pressure due to schoolwork
- Superpower Most desired superpower (Fly, Freeze time, Invisibility, Super strength, or Telepathy)
- Preference Prefers to be Famous, Happy, Healthy, or Rich

Details

The dataset gives responses for a random sample of high school seniors in Pennsylvania who participated in the Census at Schools project in 2023 and 2024.  
\*\* Updated for 4e (earlier version with seniors from 2010-2019 is now in PASeniors2019) \*\*

Source

Data from U.S. Census at School (<https://ww2.amstat.org/censusatschool/>) downloaded and used with the permission of the American Statistical Association.

---

PizzaGirl	<i>Pizza Girl Tips</i>
-----------	------------------------

---

Description

Data on tips for pizza deliveries

Format

A dataset with 24 observations on the following 2 variables.

- Tip Amount of tip (in dollars)
- Shift Data collected over three different shifts

Details

"Pizza Girl" collected data on her deliveries and tips over three different evening shifts.

Source

Pizza Girl: Statistical Analysis at <http://slice.seriousseats.com/archives/2010/04/statistical-analysis-of-a-pizza-delivery-shift-20100429.html>.

ProcessedFoods

*Ultra-Processed vs Unprocessed Food Diets***Description**

Results from an experiment comparing diets with ultra-processed food to unprocessed foods

**Format**

A data frame with 20 observations on the following 14 variables.

WeightGain Difference in weight gain

Kcal Difference in average daily caloric consumption

Sugar Difference in average daily sugar consumption in grams

Fiber Difference in average daily fiber consumption in grams

Protein Difference in average daily protein consumption in grams

Fat Difference in average daily fat consumption in grams

Carbs Difference in average daily carbohydrate consumption in grams

Ghrelin Difference in active ghrelin hormone level

Leptin Difference in leptin hormone level

PYY Difference in PYY hormone level

Glucagon Difference in glucagon hormone level

Hungry Difference in average self-reported measure of feeling hungry

Satisfied Difference in average self-reported measure of feeling satisfied

Fullness Difference in average self-reported measure of feeling full

EatingCapacity Difference in average self-reported measure of feeling hungry

**Details**

Twenty adults were admitted to the National Institutes of Health (NIH) Clinical Center and had all of their food supplied for 4 weeks. Each was given a diet of ultra-processed foods for two weeks, and a diet of unprocessed foods for two weeks, with the order of the diet randomized. The diets as presented were matched for calories, sugar, fat, fiber, and micronutrients, but then participants could choose how much of each food to eat. The cases are the 20 participants, and the variables give the difference in responses, ultra-processed response - unprocessed response, for each person.

**Source**

Hall KD, et al. (2019). "Ultra-Processed Diets Cause Excess Calorie Intake and Weight Gain: An Inpatient Randomized Controlled Trial of Ad Libitum Food Intake," *Cell Metabolism*, 30(1): 67-77. Thanks to the author for sharing the data.



PublicLibraries

*Public Library Usage***Description**

Frequency of public library use by political party

**Format**

A data frame with 2429 observations on the following 2 variables.

LibraryUse Coded as Monthly, Few Times, Rarely, or Never

Party Political party (Dem=Democrat, Ind=Independent, Rep=Republican)

**Details**

In a survey conducted by YouGov in April 2024, a sample of adults answered a question on how frequently they use public libraries and their political party affiliation

**Source**

Van Dam, A, "Who uses public libraries the most? There's a divide by religion, and politics." The Washington Post, October 2, 2024. Some values approximated from data available at <https://today.yougov.com/>.

PumpkinBeer

*Pumpkin Beer***Description**

Ratings of different kinds of pumpkin beer by a wife and husband

**Format**

A data frame with 18 observations on the following 8 variables.

Name Name of pumpkin beer

Brewer Name of brewery that produced the beer

WifeRating Rating on a 0-10 scale by the wife

HusbandRating Rating on a 0-10 scale by the husband

WifeComments Text of comments by the wife

HusbandComments Text of comments by the husband

Average Average of the two ratings (wife and husband)

Year Year the ratings were done (2011 to 2019)

**Details**

A Lock wife and husband are fans of pumpkin flavored beer, so they have each rated a variety of different brands of pumpkin beer over the years.

**Source**

Personal records

---

QuizPulse10

*Quiz vs Lecture Pulse Rates*

---

**Description**

Paired data with pulse rates in a lecture and during a quiz for 10 students

**Format**

A dataset with 10 observations on the following 3 variables.

Student	ID number for the student
Quiz	Pulse rate (beats per minute) during a quiz
Lecture	Pulse rate (beats per minute) during a lecture

**Details**

Ten students in an introductory statistics class measured their pulse rate (beats per minute) in two settings: first, in the middle of a regular class lecture and second, while taking an in-class quiz.

**Source**

In-class data collection

---

RandomP50N200

*Simulated proportions*

---

**Description**

Counts and proportions for 5000 simulated samples with  $n=200$  and  $p=0.50$

**Format**

A dataset with 5000 observations on the following two variables

Count	Number of simulated "yes" responses in 200 trials
Phat	Sample proportion ( $\text{Count}/200$ )

**Details**

Results from 5000 simulations of samples of size  $n=200$  from a population with proportion of "yes" responses at  $p=0.50$ .

**Source**

Computer simulation

---

RestaurantTips	<i>Restaurant Tips - 11 West</i>
----------------	----------------------------------

---

**Description**

Bill and tip data from the restaurant 11 West

**Format**

A dataset with 300 observations on the following 7 variables.

Bill	Size of the bill (in dollars)
Tip	Size of the tip (in dollars)
Guests	Number of people in the group
Day	Day of the week
Server	Code for specific wait person: A, B, C, D, E, F, G, or H,
Time	Lunch or Dinner
PctTip	Tip as a percentage of the bill

**Details**

Data from a sample of bills with tips during a week in March 2025 from 11 West, a restaurant in Canton, NY.

\*\* Updated for 4e (earlier version for First Crush Bistro is now RestaurantTips1e). \*\*

**Source**

Thanks to Mike Frazer and Les Baker from 11 West for providing the tipping data.

---

RetailSales	<i>Retail Sales (2009-2024)</i>
-------------	---------------------------------

---

**Description**

Monthly U.S. Retail Sales from 2009 to 2024

**Format**

A data frame with 192 observations on the following 3 variables.

Month	Month (Jan through Dec)
Year	Years from 2009 to 2024
Sales	Monthly U.S. retail sales (in billions of dollars)

**Details**

Data show the monthly retail sales (in billions) for the U.S. economy in each month from January 2009 through December 2024.

\*\* Updated for 4e (earlier versions are RetailSales2e and RetailSales2011). \*\*

**Source**

Data downloaded from <https://www.census.gov/retail/> (February 2025).

---

RockandRoll2024	<i>Rock &amp; Roll Hall of Fame (2024)</i>
-----------------	--

---

**Description**

Groups and Individuals in the Rock and Roll Hall of Fame as of 2024

**Format**

A data frame with 393 observations on the following 4 variables.

Inductee Name of the group or individual

FemaleMembers Yes if individual or member of the group is female, otherwise No

Category Type of individual or group: Early Influence, Lifetime Achievement, Non-performer, Performer, or Sideman

People Number of people in the group

**Details**

All inductees of the Rock & Roll Hall of Fame as of 2024.

\*\* Updated for 4e (earlier versions are now RockandRoll2019, RockandRoll2015, and RockandRoll2012) \*\*

**Source**

Rock & Roll Hall of Fame website, <https://www.rockhall.com/inductees/a-z>

---

RollerCoasters	<i>Roller Coasters</i>
----------------	------------------------

---

**Description**

Characteristics of a sample of roller coasters in the United States

**Format**

A data frame with 157 observations on the following 13 variables.

Coaster Name of the roller coaster

Park Name of the amusement park

City City where the coaster is located

State State where the coaster is located

Type Wood or Steel

Design Sit Down, Stand Up, Inverted, etc.

Year Year the coaster was put into service  
 Speed Top speed (in mph)  
 Height Vertical distance to the tallest point (in feet)  
 Drop Maximum vertical drop (in feet)  
 Length Total length of the ride (in feet)  
 Duration Duration of the ride (in seconds)  
 Inversions Number of inversions

### Details

Information on a sample of roller coasters at amusement parks in the US.

### Source

Roller coaster data obtained from a CODAP example at <https://codap.concord.org/>. Accessed May 2025. Much of the data came originally from the Roller Coaster Database <https://rcdb.com/>.

---

SalaryBySex	<i>Salary and Sex - 2023</i>
-------------	------------------------------

---

### Description

Salaries for college teachers (2023)

### Format

A dataset with 200 observations on the following 4 variables.

Salary	Annual salary in \$1,000's
Sex	0=female or 1=male
Age	Age in years
PhD	1=have PhD or 0=no PhD

### Details

A random sample of postsecondary teachers taken from the 2023 American Community Survey (ACS) 1-year Public Use Microdata Sample (PUMS). Updated for 4e (earlier version from 2010 is SalaryGender).

### Source

Downloaded from <https://www.census.gov/programs-surveys/acs/data/pums.html>

SampColleges

*Sample of US Post-secondary Schools***Description**

Information for a sample of 50 US post-secondary schools from the Department of Education's College Scorecard

**Format**

A data frame with 50 observations on the following 35 variables.

Name Name of the school

State State where school is located

ID ID number for school

Main Main campus? (1=yes, 0=branch campus)

Accred Accreditation agency

MainDegree Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)

HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)

Control Control of school (Private, Profit, Public)

Region Region of country (Midwest, Northeast, Southeast, Territory, West)

Locale Locale (City, Rural, Suburb, Town)

Latitude Latitude

Longitude Longitude

AdmitRate Admission rate

MidACT Median of ACT scores

AvgSAT Average combined SAT equivalent scores for admitted students

Online Only online (distance) programs

Enrollment Undergraduate enrollment

White Percent of undergraduates who report being white

Black Percent of undergraduates who report being black

Hispanic Percent of undergraduates who report being Hispanic

Asian Percent of undergraduates who report being Asian

Other Percent of undergraduates who don't report one of the above

PartTime Percent of undergraduates who are part-time students

NetPrice Average net price (cost minus aid)

Cost Average total cost for tuition, room, board, etc.

TuitionIn In-state tuition and fees

TuitionOut Out-of-state tuition and fees

TuitionFTE Net Tuition revenue per FTE student

InstructFTE Instructional spending per FTE student

FacSalary Average monthly salary for full-time faculty  
 FullTimeFac Percent of faculty that are full-time  
 Pell Percent of students receiving Pell grants  
 CompRate Completion rate (percent who finish program within 150% of normal time)  
 Debt Median debt for students who complete program  
 PctWomen Percent of women students

### Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains information from a sample of the 50 schools selected from CollegeScores. Updated for 4e (previous dataset is now SampColleges2yr3e).

### Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

---

SampColleges2yr

*Sample of College Scorecard - Two Year*

---

### Description

Information for a sample of 50 US post-secondary schools that primarily grant associate's degrees, from the Department of Education's College Scorecard

### Format

A data frame with 50 observations on the following 35 variables.

Name Name of the school  
 State State where school is located  
 ID ID number for school  
 Main Main campus? (1=yes, 0=branch campus)  
 Accred Accreditation agency  
 MainDegree Predominant undergrad degree (0=not classified, 1=certificate, 2=associate, 3=bachelors, 4=only graduate)  
 HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)  
 Control Control of school (Private, Profit, Public)  
 Region Region of country (Midwest, Northeast, Southeast, Territory, West)  
 Locale Locale (City, Rural, Suburb, Town)  
 Latitude Latitude  
 Longitude Longitude  
 AdmitRate Admission rate  
 MidACT Median of ACT scores

AvgSAT Average combined SAT equivalent scores for admitted students  
 Online Only online (distance) programs  
 Enrollment Undergraduate enrollment  
 White Percent of undergraduates who report being white  
 Black Percent of undergraduates who report being black  
 Hispanic Percent of undergraduates who report being Hispanic  
 Asian Percent of undergraduates who report being Asian  
 Other Percent of undergraduates who don't report one of the above  
 PartTime Percent of undergraduates who are part-time students  
 NetPrice Average net price (cost minus aid)  
 Cost Average total cost for tuition, room, board, etc.  
 TuitionIn In-state tuition and fees  
 TuitionOut Out-of-state tuition and fees  
 TuitionFTE Net Tuition revenue per FTE student  
 InstructFTE Instructional spending per FTE student  
 FacSalary Average monthly salary for full-time faculty  
 FullTimeFac Percent of faculty that are full-time  
 Pell Percent of students receiving Pell grants  
 CompRate Completion rate (percent who finish program within 150% of normal time)  
 Debt Median debt for students who complete program  
 PctWomen Percent of women students

### Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains information from a sample of the two-year colleges selected from all two-year colleges in CollegeScores2yr.Updated for 4e (previous dataset is now SamCollegeScores2yr3e).

### Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

---

SampColleges4yr

*Sample of College Scorecard - Four Year*

---

### Description

Information on a sample of 50 US four-year colleges and universities from the Department of Education's College Scoreboard



**Format**

A data frame with 50 observations on the following 35 variables.

Name Name of the school

State State where school is located

ID ID number for school

Main Main campus? (1=yes, 0=branch campus)

Accred Accreditation agency

MainDegree Predominant undergrad degree (3=bachelors)

HighDegree Highest degree (0=no degrees, 1=certificate, 2=associate, 3=bachelors, 4= graduate)

Control Control of school (Private, Profit, Public)

Region Region of country (Midwest, Northeast, Southeast, Territory, West)

Locale Locale (City, Rural, Suburb, Town)

Latitude Latitude

Longitude Longitude

AdmitRate Admission rate

MidACT Median of ACT scores

AvgSAT Average combined SAT equivalent scores for admitted students

Online Only online (distance) programs

Enrollment Undergraduate enrollment

White Percent of undergraduates who report being white

Black Percent of undergraduates who report being black

Hispanic Percent of undergraduates who report being Hispanic

Asian Percent of undergraduates who report being Asian

Other Percent of undergraduates who don't report one of the above

PartTime Percent of undergraduates who are part-time students

NetPrice Average net price (cost minus aid)

Cost Average total cost for tuition, room, board, etc.

TuitionIn In-state tuition and fees

TuitionOut Out-of-state tuition and fees

TuitionFTE Net Tuition revenue per FTE student

InstructFTE Instructional spending per FTE student

FacSalary Average monthly salary for full-time faculty

FullTimeFac Percent of faculty that are full-time

Pell Percent of students receiving Pell grants

CompRate Completion rate (percent who finish program within 150% of normal time)

Debt Median debt for students who complete program

PctWomen Percent of women students

## Details

The US Department of Education maintains a database through its College Scorecard project of demographic information from all active postsecondary educational institutions that participate in Title IV. This dataset contains information from a sample of the four-year colleges and universities (with full data) selected from all four-year colleges in CollegeScores4yr. Updated for 4e (previous dataset is now CollegeScores4yr3e).

## Source

Data downloaded from the US Department of Education's College Scorecard at <https://collegescorecard.ed.gov/data/> (February 2025)

---

SampCountries	<i>Sample of Countries</i>
---------------	----------------------------

---

## Description

Data on a sample of fifty countries of the world

## Format

A data frame with 50 observations on the following 29 variables.

Country Country name  
 Code Three-letter code for country  
 LandArea Size in 1000 sq. km.  
 Population Population in millions  
 Density Number of people per square kilometer  
 GDP Gross Domestic Product (in \$US) per capita  
 Rural1 Percentage of population living in rural areas  
 CO2 CO2 emissions (metric tons per capita)  
 PumpPrice Price for a liter of gasoline (\$US)  
 Military Percentage of government expenditures directed toward the military  
 Health Percentage of government expenditures directed towards healthcare  
 ArmedForces Number of active duty military personnel (in 1,000's)  
 Internet Percentage of the population with access to the internet  
 Cell Cell phone subscriptions (per 100 people)  
 HIV Percentage of the population with HIV  
 Hunger Percent of the population considered undernourished  
 Diabetes Percent of the population diagnosed with diabetes  
 BirthRate Births per 1000 people  
 DeathRate Deaths per 1000 people  
 ElderlyPop Percentage of the population at least 65 years old  
 LifeExpectancy Average life expectancy (years)

FemaleLabor Percent of females 15 - 64 in the labor force  
 Unemployment Percent of labor force unemployed  
 Renewable Percent of energy from renewable sources  
 Energy Total energy consumption (million BTU per capita)  
 Electricity Electric power consumption (kWh per capita)  
 Developed Categories for kilowatt hours per capita, 1= under 2500, 2=2500 to 5000, 3=over 5000  
 HDI Human Development Index - United Nations' measure of social and economic well being on a 0-1 scale  
 HDIGroup Categories (Very High, High, Medium, Low) based on HDI

### Details

Data from AllCountries for a random sample of 50 countries. Data from 2021-2024 depending on availability.

\*\* Updated for 4e (earlier versions are now SampCountries3e, SampCountries2e, and SampCountries1e). \*\*

### Source

Most data were gathered online from <https://data.worldbank.org/>.

Gasoline prices come from <https://tradingeconomics.com/country-list/gasoline-prices?continent=world>.

Electricity and Energy variables from U.S. Energy Information Administration, <https://www.eia.gov/international/data/world#/>

HDI variables from United Nations Human Development Report, <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI> All accessed January 2025.

---

SandP500

*S&P 500 Prices*

---

### Description

Daily data for S&P 500 Stock Index in 2024

### Format

A data frame with 252 observations on the following 6 variables.

Date Trading date (mm/dd/yyyy)

Open Opening value

High High point for the day

Low Low point for the day

Close Closing value

Volume Shares traded (in millions)

### Details

Daily prices for the S&P 500 Stock Index for trading days in 2024.

\*\* Updated for 4e (earlier versions are SandP5003e from 2018, SandP5002e from 2014, and SandP5001e from 2010). \*\*

### Source

Downloaded from <https://stooq.com>

---

SandwichAnts	<i>Sandwich Ants</i>
--------------	----------------------

---

### Description

Ant counts on samples of different sandwiches

### Format

A dataset with 24 observations on the following 5 variables.

Butter	Butter on the sandwich? no (Cases with Butter=yes are in SandwichAnts2)
Filling	Type of filling: Ham & Pickles, Peanut Butter, or Vegemite
Bread	Type of bread: Multigrain, Rye, White, or Wholemeal
Ants	Number of ants on the sandwich
Order	Trial number

### Details

As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham & pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches.

To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counted the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the factors.

This dataset has only sandwiches with no butter. The data in SandwichAnts2 adds information for samples with butter.

### Source

Margaret Mackisack, "Favourite Experiments: An Addendum to What is the Use of Experiments Conducted by Statistics Students?", Journal of Statistics Education (1994)  
<http://www.amstat.org/publications/jse/v2n1/mackisack.supp.html>

SandwichAnts2

*Sandwich Ants - Part 2***Description**

Ant counts on samples of different sandwiches

**Format**

A dataset with 48 observations on the following 5 variables.

Butter	Butter on the sandwich? no or yes
Filling	Type of filling: Ham & Pickles, Peanut Butter, or Vegemite
Bread	Type of bread: Multigrain, Rye, White, or Wholemeal
Ants	Number of ants on the sandwich
Order	Trial number

**Details**

As young students, Dominic Kelly and his friends enjoyed watching ants gather on pieces of sandwiches. Later, as a university student, Dominic decided to study this with a more formal experiment. He chose three types of sandwich fillings (vegemite, peanut butter, and ham & pickles), four types of bread (multigrain, rye, white, and wholemeal), and put butter on some of the sandwiches.

To conduct the experiment he randomly chose a sandwich, broke off a piece, and left it on the ground near an ant hill. After several minutes he placed a jar over the sandwich bit and counted the number of ants. He repeated the process, allowing time for ants to return to the hill after each trial, until he had two samples for each combination of the three factors.

**Source**

Margaret Mackisack, "Favourite Experiments: An Addendum to What is the Use of Experiments Conducted by Statistics Students?", *Journal of Statistics Education* (1994)  
<http://www.amstat.org/publications/jse/v2n1/mackisack.supp.html>

Showers

*Shower Times***Description**

Shower times and amount of water used

**Format**

A data frame with 1335 observations on the following 7 variables.

DeviceID	Identifier for a shower sensor
WaterPressure	Water pressure (in bars)
Time	Shower time (in minutes)

FlowRate Water flow rate (liters/min)  
Volume Amount of water during the shower (liters)  
Timer Was a shower timer available? (No or Yes)  
ShowerHeadType Code for the type of showerhead (2, 3, 4, 5, 6, or 8)

Details

Water usage was measured in a study at more than one hundred shower sites over a 39-week period. This dataset has results from a subset of that study (week 5) giving the time of each shower, along with water pressure, flow rate, and volume of water used. Some showers had a timer available to show users the length of time in the shower.

Source

Pereira-Doel, P., Daly, J. E. M., & Walker, I. (2024, March 22). "Beyond the water flow rate: water pressure and smart timers impact shower efficiency" <https://doi.org/10.17605/OSF.IO/NTZGC>

---

SkateboardPrices	Skateboard Prices in (2025)
------------------	-----------------------------

---

Description

Prices of skateboards for sale online in 2025

Format

A dataset with 20 observations on the following variable.

Price Selling price in dollars (including delivery)

Details

Prices (including delivery) for skateboards offered for sale on eBay.  
\*\*Updated for 4e (earlier version is nowSkateboardPrices2012) \*\*

Source

Random sample taken from skateboards available for sale on eBay in March 2025.

SleepCaffeine

*Sleep Caffeine***Description**

Experiment to compare word recall after sleep or caffeine

**Format**

A dataset with 24 observations on the following 2 variables.

Group	Treatment: Caffeine or Sleep
Words	Number of words recalled

**Details**

A random sample of 24 adults were divided equally into two groups and given a list of 24 words to memorize. During a break, one group takes a 90 minute nap while another group is given a caffeine pill. The response variable is the number of words participants are able to recall following the break.

**Source**

Mednick, Cai, Kanady, and Drummond, "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory", Behavioural Brain Research, 193 (2008), 79-86.

SleepStudy

*Sleep Study***Description**

Data from a study of sleep patterns for college students.

**Format**

A dataset with 253 observations on the following 27 variables.

Gender	1=male, 0=female
ClassYear	Year in school, 1=first year, ..., 4=senior
LarkOwl	Early riser or night owl? Lark, Neither, or Owl
NumEarlyClass	Number of classes per week before 9 am
EarlyClass	Indicator for any early classes
GPA	Grade point average (0-4 scale)
ClassesMissed	Number of classes missed in a semester
CognitionZscore	Z-score on a test of cognitive skills
PoorSleepQuality	Measure of sleep quality (higher values are poorer sleep)
DepressionScore	Measure of degree of depression
AnxietyScore	Measure of amount of anxiety

StressScore	Measure of amount of stress
DepressionStatus	Coded depression score: normal, moderate, or severe
AnxietyStatus	Coded anxiety score: normal, moderate, or severe
Stress	Coded stress score: normal or high
DASScore	Combined score for depression, anxiety and stress
Happiness	Measure of degree of happiness
AlcoholUse	Self-reported: Abstain, Light, Moderate, or Heavy
Drinks	Number of alcoholic drinks per week
WeekdayBed	Average weekday bedtime (24.0=midnight)
WeekdayRise	Average weekday rise time (8.0=8 am)
WeekdaySleep	Average hours of sleep on weekdays
WeekendBed	Average weekend bedtime (24.0=midnight)
WeekendRise	Average weekend rise time (8.0=8 am)
WeekendSleep	Average weekend bedtime (24.0=midnight)
AverageSleep	Average hours of sleep for all days
AllNighter	Had an all-nighter this semester? 1=yes, 0=no

### Details

The data were obtained from a sample of students who did skills tests to measure cognitive function, completed a survey that asked many questions about attitudes and habits, and kept a sleep diary to record time and quality of sleep over a two week period.

### Source

Onyper, S., Thacher, P., Gilbert, J., Gradess, S., "Class Start Times, Sleep, and Academic Performance in College: A Path Analysis," April 2012; 29(3): 318-335. Thanks to the authors for supplying the data.

---

Smiles

*Smiles*

---

### Description

Experiment to study effect of smiling on leniency in judicial matters

### Format

A dataset with 68 observations on the following 2 variables.

Leniency	Score assigned by a judgment panel (higher is more lenient)
Group	Treatment group: neutral or smile

### Details

Hecht and LeFrance conducted a study examining the effect of a smile on the leniency of disciplinary action for wrongdoers. Participants in the experiment took on the role of members of a college disciplinary panel judging students accused of cheating. For each suspect, along with a description of the offense, a picture was provided with either a smile or neutral facial expression. A leniency score was calculated based on the disciplinary decisions made by the participants.



**Source**

LaFrance, M., & Hecht, M. A., "Why smiles generate leniency", *Personality and Social Psychology Bulletin*, 21, 1995, 207-214.

SpeedDating

*Speed Dating***Description**

Data from a sample of four minute speed dates.

**Format**

A dataset with 276 observations on the following 22 variables.

DecisionM	Would the male like another date? 1=yes 0=no
DecisionF	Would the female like another date? 1=yes 0=no
LikeM	How much the male likes his partner (1-10 scale)
LikeF	How much the female likes her partner (1-10 scale)
PartnerYesM	Male's estimate of chance the female wants another date (1-10 scale)
PartnerYesF	Female's estimate of chance the male wants another date (1-10 scale)
AgeM	Male's age (in years)
AgeF	Females age (in years)
RaceM	Male's race: Asian Black Caucasian Latino Other
RaceF	Female's race: Asian Black Caucasian Latino Other
AttractiveM	Male's rating of female's attractiveness (1-10 scale)
AttractiveF	Female's rating of male's attractiveness (1-10 scale)
SincereM	Male's rating of female's sincerity (1-10 scale)
SincereF	Female's rating of male's sincerity (1-10 scale)
IntelligentM	Male's rating of female's intelligence (1-10 scale)
IntelligentF	Female's rating of male's intelligence (1-10 scale)
FunM	Male's rating of female as fun (1-10 scale)
FunF	Female's rating of male as fun (1-10 scale)
AmbitiousM	Male's rating of female's ambition (1-10 scale)
AmbitiousF	Female's rating of male's ambition (1-10 scale)
SharedInterestsM	Male's rating of female's shared interests (1-10 scale)
SharedInterestsF	Female's rating of male's shared interests (1-10 scale)

**Details**

Participants were students at Columbia's graduate and professional schools, recruited by mass email, posted fliers, and fliers handed out by research assistants. Each participant attended one speed dating session, in which they met with each participant of the opposite sex for four minutes. Order and session assignments were randomly determined. After each four minute "speed date," participants filled out a form rating their date on a scale of 1-10 on various attributes. Only data from the first date in each session is recorded here.

**Source**

Gelman, A. and Hill, J., *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press: New York, 2007

SplitBill

*Split Bill vs Individual Meal Costs***Description**

Meal costs when ordering individually vs splitting a bill

**Format**

A dataset with 48 observations on the following 4 variables.

Payment	Payment method: Individual or Split
Sex	F = female or M = male
Items	Number of items ordered
Cost	Cost of items ordered in Israeli new shekel's (ILS)

**Details**

Subjects were 48 Israeli students who were randomly assigned to eat in groups of six (three males and three females) at a restaurant. Half the groups were told that they would pay for meals individually and half were told that the group would split the bill equally. The number of items ordered and cost (in Israeli new shekels) was recorded for each individual.

**Source**

Gneezy, U., Haruvy, E., and Yafe, H. "The Inefficiency of Splitting the Bill," The Economic Journal, 2004; 114, 265-280.

StatGrades

*Statistics Exam Grades***Description**

Grades on statistics exams

**Format**

A dataset with 50 observations on the following 3 variables.

Exam1	Score (out of 100 points) on the first exam
Exam2	Score (out of 100 points) on the second exam
Final	Score (out of 100 points) on the final exam

**Details**

Exam scores for a sample of students who completed a course using Statistics: Unlocking the Power of Data as a text. The dataset contains scores on Exam1 (Chapters 1 to 4), Exam2 (Chapters 5 to 8), and the Final exam (entire book).

Source

Random selection of students in an introductory statistics course.

---

StockChanges	<i>Stock Changes (2024)</i>
--------------	-----------------------------

---

Description

Stock price change for a sample of stocks from the S&P 500 (December 2-6, 2024)

Format

A dataset with 50 observations on the following 2 variables.

Symbol Ticker symbol for the stock  
SPChange Change in stock price (in dollars)

Details

A random sample of 50 companies from Standard & Poor’s index of 500 companies was selected. The change in the price of the stock (in dollars) over the 5-day period from December 2 - 6, 2024 was recorded for each company in the sample.

Source

Data obtained from Kaggle at <https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks>

---

StorySpoilers	<i>Story Spoilers</i>
---------------	-----------------------

---

Description

Ratings for stories with and without spoilers

Format

A dataset with 12 observations on the following 3 variables.

Story	ID for story
Spoiler	Average (0-10) rating for spoiler version
Original	Average (0-10) rating for original version

Details

This study investigated whether a story spoiler that gives away the ending early diminishes suspense and hurts enjoyment. For twelve different short stories, the study’s authors created a second version in which a spoiler paragraph at the beginning discussed the story and revealed the outcome. Each version of the twelve stories was read by at least 30 people and rated on a 1 to 10 scale to create an overall rating for the story, with higher ratings indicating greater enjoyment of the story. Stories 1 to 4 were ironic twist stories, stories 5 to 8 were mysteries, and stories 9 to 12 were literary stories.

**Source**

Leavitt, J. and Christenfeld, N., "Story Spoilers Don't Spoil Stories," Psychological Science, published OnlineFirst, August 12, 2011.

---

StressedMice	<i>Stressed Mice</i>
--------------	----------------------

---

**Description**

Time in darkness for mice in different environments

**Format**

A dataset with 14 observations on the following 2 variables.

Time	Time spent in darkness (in seconds)
Environment	Type of environment: Enriched or Standard

**Details**

In the study, mice were randomly assigned to either an enriched environment where there was an exercise wheel available, or a standard environment with no exercise options. After three weeks in the specified environment, for five minutes a day for two weeks, the mice were each exposed to a "mouse bully" - a mouse who was very strong, aggressive, and territorial. One measure of mouse anxiety is amount of time hiding in a dark compartment, with mice who are more anxious spending more time in darkness. The amount of time spent in darkness is recorded for each of the mice.

**Source**

Data approximated from summary statistics in: Lehmann and Herkenham, "Environmental Enrichment Confers Stress Resiliency to Social Defeat through an Infralimbic Cortex-Dependent Neuroanatomical Pathway", The Journal of Neuroscience, April 20, 2011, 31(16):61596173.

---

StudentSurvey	<i>Student Survey Data</i>
---------------	----------------------------

---

**Description**

Data from a survey of students in introductory statistics courses

**Format**

A data frame with 362 observations on the following 17 variables.

Year Year in school

Sex F=female or M=male

Smoke Smoker? No or Yes

Award Preferred award: Academy, Nobel, or Olympic

HigherSAT Which SAT is higher? Math or Verbal

Exercise Hours of exercise per week

TV Hours of TV viewing per week

Height Height (in inches)

Weight Weight (in pounds)

Siblings Number of siblings

BirthOrder Birth order, 1=oldest

VerbalSAT Verbal SAT score

MathSAT Math SAT scorer

SAT Combined Verbal + Math SAT

GPA College grade point average

Pulse Pulse rate (beats per minute)

Piercings Number of body piercings

**Details**

Data from an in-class survey given to introductory statistics students over several years. Note the Sex variable was labeled as Gender in earlier versions of this dataset. We acknowledge that this binary dichotomization is not a complete or inclusive representation of reality.

**Source**

In-class student survey

---

SynchronizedMovement    *Synchronized Movement*

---

**Description**

Effects of synchronized movement activities

**Format**

A dataset with 264 observations on the following 11 variables.

Sex	f = female or m = male
Group	Type of activity. Coded as HS+HE, HS+LE, LS+HE, or LS+LE for High/Low Synchronization + High/Low Exertion
Synch	Synchronized activity? yes or no
Exertion	Exertion level: high or low
PainToleranceBefore	Measure of pain tolerance (mm Hg) before activity
PainTolerance	Measure of pain tolerance (mm Hg) after activity
PainTolDiff	Difference (after - before) in pain tolerance
MaxPressure	Reached the maximum pressure (300 mm Hg) when testing pain tolerance (after)
CloseBefore	Rating of closeness to the group before activity (1=least close to 7=most close)
CloseAfter	Rating of closeness to the group after activity (1=least close to 7=most close)
CloseDiff	Change on closeness rating (after - before)

**Details**

From a study of 264 high school students in Brazil to examine the effect of doing synchronized movements (such as marching in step or doing synchronized dance steps) and the effect of exertion on variables, such as pain tolerance and attitudes towards others. Students were randomly assigned to activities that involved synchronized or non-synchronized movements involving high or low levels of exertion. Pain tolerance was measured with a blood pressure cuff, going to a maximum possible reading of 300 mmHg.

**Source**

Tarr B, Launay J, Cohen E, and Dunbar R, "Synchrony and exertion during dance independently raise pain threshold and encourage social bonding," *Biology Letters*, 11(10), October 2015.

---

TenCountries

*Ten Countries*

---

**Description**

A subset of the AllCountries data for a random sample of ten countries

**Format**

A data frame with 10 observations on the following 4 variables.

Country	Country name
Code	Three-letter country code
Area	Size in 1000 sq. kilometers
PctRural	Percentage of population living in rural areas

**Details**

Area and percent rural for a sample of ten countries from AllCountries dataset.

\*\* Updated for 4e (earlier versions are now Ten Countries3e, TenCountries2e, and TenCountries1e)  
\*\*

**Source**

Data collected from the World Bank website, <https://www.worldbank.org/en/home>

---

TextbookCosts	<i>Textbook Costs (2025)</i>
---------------	------------------------------

---

**Description**

Prices for textbooks for different courses in Spring 2025

**Format**

A data frame with 40 observations on the following 3 variables.

Field General discipline of the course: Arts, Humanities, NaturalScience, or SocialScience

Books Number of books required

Cost Total cost (in dollars) for required books

**Details**

Data are from samples of ten courses in each of four disciplines at a liberal arts college. For each course the bookstore's website lists the required texts(s) and costs (for new books). Data were collected for the Spring 2025 semester.

**Source**

Bookstore online site

---

ToenailArsenic	<i>Toenail Arsenic</i>
----------------	------------------------

---

**Description**

Arsenic in toenails of 19 people using private wells in New Hampshire

**Format**

A dataset with 19 observations on the following variable.

Arsenic Level of arsenic found in toenails (ppm)

**Details**

Level of arsenic was measured in toenails of 19 subjects from New Hampshire, all with private wells as their main water source.

**Source**

Adapted from Karagas, et.al., "Toenail Samples as an Indicator of Drinking Water Arsenic Exposure", Cancer Epidemiology, Biomarkers and Prevention 1996;5:849-852.

---

TrafficFlow	<i>Traffic Flow</i>
-------------	---------------------

---

**Description**

Traffic flow times from a simulation with timed and flexible traffic lights

**Format**

A dataset with 24 observations on the following 3 variables.

Timed	Delay time (in minutes) for fixed timed lights
Flexible	Delay time (in minutes) for flexible communicating lights
Difference	Difference (Timed-Flexible) for each simulation

**Details**

Engineers in Dresden, Germany were looking at ways to improve traffic flow by enabling traffic lights to communicate information about traffic flow with nearby traffic lights. The data show results of one experiment where they simulated buses moving along a street and recorded the delay time (in seconds) for both a fixed time and a flexible system of lights. The process was repeated under both conditions for a sample of 24 simulated scenarios.

**Source**

Lammer and Helbing, "Self-Stabilizing decentralized signal control of realistic, saturated network traffic", Santa Fe Institute working paper 10-09-019, September 2010.

---

USStates	<i>US State Data</i>
----------	----------------------

---

**Description**

Various data for all 50 US States.

**Format**

A data frame with 50 observations on the following 22 variables.

State	State name
HouseholdIncome	Median household income (in \$1,000's)
Region	MW=Midwest, NE=Northeast, S=South, W=West
Population	Number of residents (in millions for 2023)
EighthGradeMath	Average score NAEP mathematics for 8th-grade students (2024)
HighSchool	% of residents (ages 25-34) who are high school graduates
College	% of residents (ages 25-34) who are college graduates
IQ	Estimated mean IQ score of residents



GSP Gross state product (in \$1,000's per capita)  
 Vegetables % of residents eating vegetables at least once per day  
 Fruit % of residents eating fruit at least once per day  
 Smokers % of residents who smoke  
 PhysicalActivity % who do 150+ minutes of aerobic physical activity per week  
 Obese % obese residents (BMI 30+)  
 NonWhite % nonwhite or Hispanic residents  
 HeavyDrinkers % heavy drinkers ( men: 14+ drinks/week, women 7+ drinks/week)  
 Electoral Number of state votes in the presidential electoral college  
 BidenVote Proportion of votes for Democrat Joe Biden in 2020 presidential election  
 Elect2020 State winner in 2020 presidential election (D=Biden, R=Trump, DR=Split)  
 TwoParents % of children living in two-parent households  
 StudentSpending School spending (in \$1,000 per pupil)  
 Insured % of adults (ages 19-64) who have any kind of health coverage  
 Poverty % of families with income below the poverty line

## Details

Information from each of the 50 states of the United States. Years vary from 2021 to 2024 depending on data availability.

\*\* Updated for 4e (earlier versions are now USStates3e, USStates2e, and USStates1e) \*\*

## Source

U.S. Census Bureau, American Community Survey <https://data.census.gov/all?q=ACS>.

Table DP03 - HouseholdIncome, Insured, Poverty

Table B01003 - Population

Table C23008 - TwoParents

Table S1501 - HighSchool, College

World Population Review <https://worldpopulationreview.com/state-rankings/> for IQ, Smokers, and GSP.

2020 Election results <https://www.fec.gov/resources/cms-content/documents/federalelections2020.pdf> for Electoral, BidenVote, and Elect2020.

National Assessment of Educational Progress (NAEP) <https://www.nationsreportcard.gov/profiles/stateprofile?sfj=NP&chort=2&sub=MAT&sj=&st=MN&year=2024R3> for EighthGradeMath.

National Center for EducationStatistics (NCES) [https://nces.ed.gov/programs/digest/d23/tables/dt23\\_236.65.asp?current=yes](https://nces.ed.gov/programs/digest/d23/tables/dt23_236.65.asp?current=yes) for StudentSpending

Behavioral Risk Factors Surveillance System (BRFSS) <https://www.cdc.gov/brfss/brfssprevalence/> for Fruit, Vegetables, PhysicalActivity, HeavyDrinkers, NonWhite, and Obese.

---

 WaterStriders

*Water Striders*


---

### Description

Mating activity for water striders

### Format

A dataset with 10 observations on the following 3 variables.

AggressiveMale	Hyper-aggressive male in group? No or Yes
FemalesHiding	Proportion of time the female water striders were in hiding
MatingActivity	Measure of mean mating activity (higher numbers meaning more mating)

### Details

Water striders are common bugs that skate across the surface of water. Water striders have different personalities and some of the males are hyper-aggressive, meaning they jump on and wrestle with any other water strider near them. Individually, because hyper-aggressive males are much more active, they tend to have better mating success than more inactive striders. This study examined the effect they have on a group. Four males and three females were put in each of ten pools of water. Half of the groups had a hyper-aggressive male as one of the males and half did not. The proportion of time females are in hiding was measured for each of the 10 groups, and a measure of mean mating activity was also measured with higher numbers meaning more mating.

### Source

Sih, A. and Watters, J., "The mix matters: behavioural types and group dynamics in water striders," *Behaviour*, 2005; 142(9-10): 1423.

---

 WaterTaste

*WaterTaste*


---

### Description

Blind taste test to compare brands of bottled water

### Format

A dataset with 100 observations on the following 10 variables.

Gender	Gender of respondent: F=Female M=Male
Age	Age (in years)
Class	Year in school F=First year J=Junior O=Other P S0=Sophomore SR=Senior
UsuallyDrink	Usual source of drinking water: Bottled, Filtered, or Tap
FavBotWatBrand	Favorite brand of bottled water
Preference	Order of preference: A=Sams Choice, B=Aquafina, C=Fiji, and D=Tap water

First	Top choice among Aquafina, Fiji, SamsChoice, or Tap
Second	Second choice
Third	Third choice
Fourth	Fourth choice

### Details

Result from a blind taste test comparing four different types of water (Sam's Choice, Aquafina, Fiji, and tap water). Participants rank ordered waters when presented in a random order.

### Source

"Water Taste Test Data" by M. Leigh Lunsford and Alix D. Dowling Finch in the Journal of Statistics Education (Vol 18, No, 1) 2010  
<http://www.amstat.org/publications/jse/v18n1/lunsford.pdf>

---

Wetsuits	<i>Wetsuits</i>
----------	-----------------

---

### Description

Swim velocity (for 1500 meters) with and without wearing a wetsuit

### Format

A dataset with 12 observations on the following 4 variables.

Wetsuit	Maximum swim velocity (m/sec) when wearing a wetsuit
NoWetsuit	Maximum swim velocity (m/sec) when wearing a regular bathing suit
Gender	Gender of swimmer: F or M
Type	Type of athlete: swimmer or triathlete

### Details

A study tested whether wearing wetsuits influences swimming velocity. Twelve competitive swimmers and triathletes swam 1500m at maximum speed twice each; once wearing a wetsuit and once wearing a regular bathing suit. The order of the trials was randomized. Each time, the maximum velocity in meters/sec of the swimmer was recorded.

### Source

de Lucas, R.D., Balildan, P., Neiva, C.M., Greco, C.C., Denadai, B.S. (2000). "The effects of wetsuits on physiological and biomechanical indices during swimming," Journal of Science and Medicine in Sport, 3 (1): 1-8.

---

XylitolEffect	<i>Xylitol vs. Sugar - Blood Clotting</i>
---------------	---

---

**Description**

Blood clotting measures with Xylitol and Sugar

**Format**

A data frame with 10 observations on the following 4 variables.

- Subject Subject ID number
- Xylitol Blood clotting measure after taking xylitol
- Sugar Blood clotting measure after taking sugar
- Diff Difference=Xylitol-Sugar

**Details**

Researchers were interested in how xylitol might affect blood clotting. Ten subjects had a measure of blood clotting taken 30 minutes after ingesting a drink sweetened with xylitol (a sugar substitute) and again after a sugar-sweetened drink.

**Source**

Witkowski M, et al., “Xylitol is prothrombotic and associated with cardiovascular risk,” European Heart Journal, 45 (27), July 14, 2024.  
Values are estimated from information given in the paper

---

YoungBlood	<i>Young Blood</i>
------------	--------------------

---

**Description**

Effects of transfusions of young blood on exercise endurance in mice

**Format**

A dataset with 30 observations on the following 2 variables.

- Plasma Whether the blood came from a Young or Old mouse
- Runtime Maximum treadmill run time (in minutes) in a 90-minute window

**Details**

The data come from a study to see if transfusions of blood plasma from young mice (equivalent to about a 25-year-old person) can counteract or reverse brain aging in old mice (equivalent to about a 70-year-old person.) Old mice were randomly assigned to receive plasma from either a young mice or another old mouse, and exercise endurance was measured.

**Source**

Data come from two references, and are estimated from summary statistics and graphs.

Sanders L, "Young blood proven good for old brain," Science News, 185(11), May 31, 2014.

Manisha S, et al., "Restoring Systemic GDF11 Levels Reverses Age-Related Dysfunction in Mouse Skeletal Muscle," Science, 9 May 2014.

# Index

## \* datasets

ACS, [4](#)  
AllCountries, [5](#)  
APMultipleChoice, [6](#)  
April14Temps, [7](#)  
BaseballHits2024, [7](#)  
BaseballSalaries2024, [8](#)  
BaseballTimes, [9](#)  
Benford, [9](#)  
BikeCommute, [10](#)  
BMI, [11](#)  
BodyFat, [11](#)  
BodyTemp50, [12](#)  
BootAtlantaCorr, [13](#)  
CaffeineTaps, [13](#)  
CAOSExam, [14](#)  
CarbonDioxide, [14](#)  
CarDepreciation, [15](#)  
Cars2025, [15](#)  
CarsEV2025, [16](#)  
Celtics2024, [17](#)  
Cereal, [19](#)  
CityTemps, [19](#)  
CocaineTreatment, [20](#)  
ColaCalcium, [20](#)  
CollegeExperience, [21](#)  
CollegeScores, [22](#)  
CollegeScores2yr, [24](#)  
CollegeScores4yr, [25](#)  
CommuteAtlanta, [27](#)  
CommuteStLouis, [27](#)  
CompassionateRats, [28](#)  
COVIDCollege, [28](#)  
COVIDCollegeStacked, [30](#)  
CricketChirps, [31](#)  
DDS, [32](#)  
DecemberFlights, [33](#)  
DietDepression, [33](#)  
Digits, [34](#)  
DrugResistance, [35](#)  
EducationLiteracy, [35](#)  
ElectionMargin, [36](#)  
EmployedACS, [37](#)

ExerciseHours, [37](#)  
FacebookFriends, [38](#)  
FireAnts, [39](#)  
FisherIris, [39](#)  
FishGills12, [40](#)  
FishGills3, [40](#)  
Flight475, [41](#)  
FloridaLakes, [42](#)  
FootballBrain, [42](#)  
ForestFires, [43](#)  
GeneticDiversity, [44](#)  
GlobalInternet2024, [45](#)  
GolfRound, [45](#)  
GPABySex, [46](#)  
GroundhogDay, [46](#)  
HappyPlanetIndex, [47](#)  
HeatCognition, [48](#)  
HeightData, [48](#)  
HockeyPenalties2025, [49](#)  
HollywoodMovies, [50](#)  
HomesForSale, [51](#)  
HomesForSaleCA, [51](#)  
HomesForSaleCanton, [52](#)  
HomesForSaleNY, [52](#)  
HomingPigeons, [53](#)  
Honeybee2023, [54](#)  
HoneybeeCircuits, [54](#)  
HoneybeeWaggle, [55](#)  
HotDogs2024, [55](#)  
HouseStarts2024, [56](#)  
HumanTears25, [56](#)  
HumanTears50, [57](#)  
Hurricanes2024, [58](#)  
ICUAdmissions, [58](#)  
ImmuneTea, [59](#)  
InkjetPrinters, [60](#)  
LifeExpectancyVehicles, [60](#)  
LightatNight, [61](#)  
LightatNight4Weeks, [62](#)  
LightatNight8Weeks, [62](#)  
MalevolentUniformsNFL, [63](#)  
MalevolentUniformsNHL, [64](#)  
MammalLongevity, [64](#)

- ManhattanApartments, 65
- MarriageAges, 66
- MastersGolf, 66
- MateChoice, 67
- MentalMuscle, 67
- MindsetMatters, 68
- MustangPrice, 69
- NBAPlayers2024, 69
- NBAStandings2024, 70
- NFLContracts2024, 71
- NFLPreseason2024, 72
- NFLScores2024, 72
- NHANES, 73
- Ninja, 74
- NutritionStudy, 75
- OlympicMarathon2024, 76
- OrganicEffect, 76
- OttawaSenators2024, 77
- PASeniors, 78
- PizzaGirl, 79
- ProcessedFoods, 80
- PublicLibraries, 81
- PumpkinBeer, 81
- QuizPulse10, 82
- RandomP50N200, 82
- RestaurantTips, 83
- RetailSales, 83
- RockandRoll2024, 84
- RollerCoasters, 84
- SalaryBySex, 85
- SampColleges, 86
- SampColleges2yr, 87
- SampColleges4yr, 88
- SampCountries, 90
- SandP500, 91
- SandwichAnts, 92
- SandwichAnts2, 93
- Showers, 93
- SkateboardPrices, 94
- SleepCaffeine, 95
- SleepStudy, 95
- Smiles, 96
- SpeedDating, 97
- SplitBill, 98
- StatGrades, 98
- StockChanges, 99
- StorySpoilers, 99
- StressedMice, 100
- StudentSurvey, 100
- SynchronizedMovement, 101
- TenCountries, 102
- TextbookCosts, 103
- ToenailArsenic, 103
- TrafficFlow, 104
- USStates, 104
- WaterStriders, 106
- WaterTaste, 106
- Wetsuits, 107
- XylitolEffect, 108
- YoungBlood, 108
- ACS, 4
- AllCountries, 5
- APMultipleChoice, 6
- April14Temps, 7
- BaseballHits2024, 7
- BaseballSalaries2024, 8
- BaseballTimes, 9
- Benford, 9
- BikeCommute, 10
- BMI, 11
- BodyFat, 11
- BodyTemp50, 12
- BootAtlantaCorr, 13
- CaffeineTaps, 13
- CAOSExam, 14
- CarbonDioxide, 14
- CarDepreciation, 15
- Cars2025, 15
- CarsEV2025, 16
- Celtics2024, 17
- Cereal, 19
- CityTemps, 19
- CocaineTreatment, 20
- ColaCalcium, 20
- CollegeExperience, 21
- CollegeScores, 22
- CollegeScores2yr, 24
- CollegeScores4yr, 25
- CommuteAtlanta, 27
- CommuteStLouis, 27
- CompassionateRats, 28
- COVIDCollege, 28
- COVIDCollegeStacked, 30
- CricketChirps, 31
- DDS, 32
- DecemberFlights, 33
- DietDepression, 33
- Digits, 34
- DrugResistance, 35
- EducationLiteracy, 35

- ElectionMargin, [36](#)
- EmployedACS, [37](#)
- ExerciseHours, [37](#)
- FacebookFriends, [38](#)
- FireAnts, [39](#)
- FisherIris, [39](#)
- FishGills12, [40](#)
- FishGills3, [40](#)
- Flight475, [41](#)
- FloridaLakes, [42](#)
- FootballBrain, [42](#)
- ForestFires, [43](#)
- GeneticDiversity, [44](#)
- GlobalInternet2024, [45](#)
- GolfRound, [45](#)
- GPAbySex, [46](#)
- GroundhogDay, [46](#)
- HappyPlanetIndex, [47](#)
- HeatCognition, [48](#)
- HeightData, [48](#)
- HockeyPenalties2025, [49](#)
- HollywoodMovies, [50](#)
- HomesForSale, [51](#)
- HomesForSaleCA, [51](#)
- HomesForSaleCanton, [52](#)
- HomesForSaleNY, [52](#)
- HomingPigeons, [53](#)
- Honeybee2023, [54](#)
- HoneybeeCircuits, [54](#)
- HoneybeeWaggle, [55](#)
- HotDogs2024, [55](#)
- HouseStarts2024, [56](#)
- HumanTears25, [56](#)
- HumanTears50, [57](#)
- Hurricanes2024, [58](#)
- ICUAdmissions, [58](#)
- ImmuneTea, [59](#)
- InkjetPrinters, [60](#)
- LifeExpectancyVehicles, [60](#)
- LightatNight, [61](#)
- LightatNight4Weeks, [62](#)
- LightatNight8Weeks, [62](#)
- MalevolentUniformsNFL, [63](#)
- MalevolentUniformsNHL, [64](#)
- MammalLongevity, [64](#)
- ManhattanApartments, [65](#)
- MarriageAges, [66](#)
- MastersGolf, [66](#)
- MateChoice, [67](#)
- MentalMuscle, [67](#)
- MindsetMatters, [68](#)
- MustangPrice, [69](#)
- NBAPlayers2024, [69](#)
- NBAStandings2024, [70](#)
- NFLContracts2024, [71](#)
- NFLPreseason2024, [72](#)
- NFLScores2024, [72](#)
- NHANES, [73](#)
- Ninja, [74](#)
- NutritionStudy, [75](#)
- OlympicMarathon2024, [76](#)
- OrganicEffect, [76](#)
- OttawaSenators2024, [77](#)
- PASeniors, [78](#)
- PizzaGirl, [79](#)
- ProcessedFoods, [80](#)
- PublicLibraries, [81](#)
- PumpkinBeer, [81](#)
- QuizPulse10, [82](#)
- RandomP50N200, [82](#)
- RestaurantTips, [83](#)
- RetailSales, [83](#)
- RockandRoll2024, [84](#)
- RollerCoasters, [84](#)
- SalaryBySex, [85](#)
- SampColleges, [86](#)
- SampColleges2yr, [87](#)
- SampColleges4yr, [88](#)
- SampCountries, [90](#)
- SandP500, [91](#)
- SandwichAnts, [92](#)
- SandwichAnts2, [93](#)
- Showers, [93](#)
- SkateboardPrices, [94](#)
- SleepCaffeine, [95](#)
- SleepStudy, [95](#)
- Smiles, [96](#)
- SpeedDating, [97](#)
- SplitBill, [98](#)
- StatGrades, [98](#)
- StockChanges, [99](#)
- StorySpoilers, [99](#)
- StressedMice, [100](#)
- StudentSurvey, [100](#)
- SynchronizedMovement, [101](#)



TenCountries, [102](#)  
TextbookCosts, [103](#)  
ToenailArsenic, [103](#)  
TrafficFlow, [104](#)  
  
USStates, [104](#)  
  
WaterStriders, [106](#)  
WaterTaste, [106](#)  
Wetsuits, [107](#)  
  
XylitolEffect, [108](#)  
  
YoungBlood, [108](#)