

COMPUTATIONAL JOURNALISM

Tuesday Evenings, Spring of 2023
Instructors, Mark Hansen and Wayne Lee

STATISTICIAN'S BRIEFING



You!

What experience do you have working with data, in or out of journalism?



Chang, Duojiao
dc3667



Cruz-Luviano, Alma
alc2313



Fan, Hangyu
hf2465



Galloway, Alana
amg2425



Jiang, Lai
lj2562



Marx, Sarah
sjm2255



Patwardhan, Anvita
ap4302



Shailer, Daniel
dls2226



Woods, Ellen
ehw2134



Wu, Ruoheng
rw2958



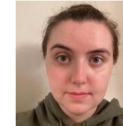
Brahmabhatt, Manan
mb4984



Fan, Yongyang
yf2635



Gu, Jingyu
jg4536



Hanger, Rachel
rah2240



Hawkins, Diahmin
dlh2166



Lam, Daniel
dr12163



Li, Chongzhi
cl4324



Nguyen, Kieu-Giang
kn2521



Ngwenya, Mengezi
mn3040



Wei, Jiahao
jw4312



Zhou, Hongyan
hz2827

Statistics making an impact

John Pullinger

House of Commons Library, London, UK

[The address of the President, delivered to The Royal Statistical Society on Wednesday, June 26th, 2013]

Summary. Statistics provides a special kind of understanding that enables well-informed decisions. As citizens and consumers we are faced with an array of choices. Statistics can help us to choose well. Our statistical brains need to be nurtured: we can all learn and practise some simple rules of statistical thinking. To understand how statistics can play a bigger part in our lives today we can draw inspiration from the founders of the Royal Statistical Society. Although in today's world the information landscape is confused, there is an opportunity for statistics that is there to be seized. This calls for us to celebrate the discipline of statistics, to show confidence in our profession, to use statistics in the public interest and to champion statistical education. The Royal Statistical Society has a vital role to play.

Keywords: Chartered Statistician; Citizenship; Economic growth; Evidence; getstats; Justice; Open data; Public good; The state; Wise choices

1. Introduction

Dictionarys trace the source of the word statistics from the Latin ‘status’, the state, to the Italian ‘statista’, one skilled in statecraft, and on to the German ‘Statistik’, the science dealing with data about the condition of a state or community. The *Oxford English Dictionary* brings ‘statistics’ into English in 1787. Florence Nightingale held that ‘the thoughts and purpose of the Deity are only to be discovered by the statistical study of natural phenomena... the application of the results of such study [is] the religious duty of man’ (Pearson, 1924).

Statistics is about matters of the highest importance in human affairs. It is about our comprehension of the world around us and how that comprehension affects the lives we live: our health, wealth and wellbeing; our understanding of the state of our planet and the condition of our local communities; our choices at work and at play.

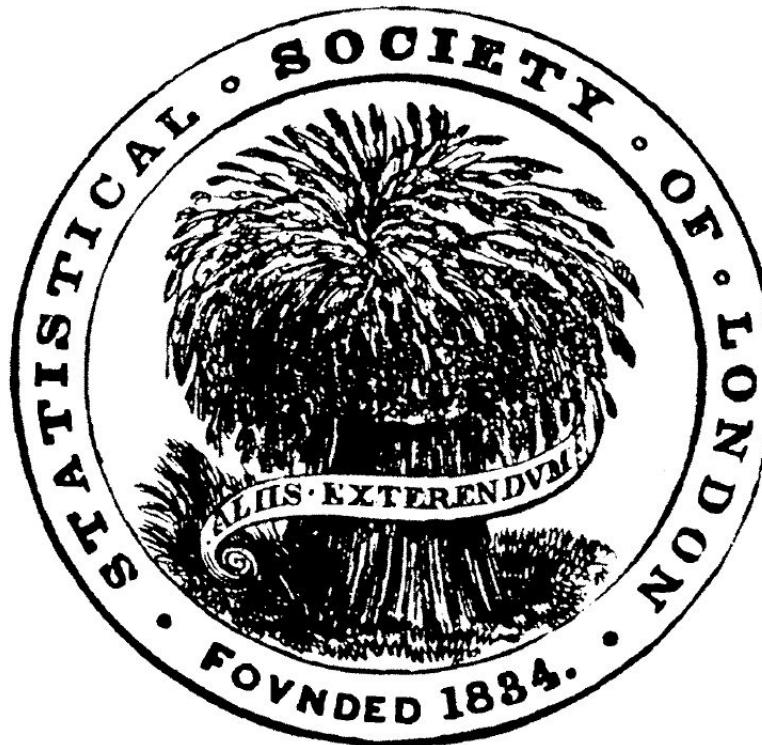
In this address I shall explore the interface between statistics and decision making. I shall identify the principles which give statistics its extraordinary ability to shape our thinking about how we view the world. I shall highlight the leading role played by the Royal Statistical Society in developing the discipline and the profession of statistics since 1834 and draw attention to why it matters so much at this time.

I shall conclude with a call to action for statistics to give people the power to make good choices. A call that chimes with the objectives of this, the International Year of Statistics (see <http://www.statistics2013.org/about-us/>), which is being promoted by the American Statistical Association, the Institute of Mathematical Statistics, the International Biometric Society and the International Statistical Institute (and the Bernoulli Society) as well as the Royal Statistical Society along with over 2000 other organizations around the world:

Address for correspondence: John Pullinger, House of Commons Library, London, SW1A 0AA, UK.
E-mail: j.pullinger@rss.org.uk

© 2013 Royal Statistical Society

0964–1998/13/176819



Principles

Anyone seems to be able to come up with a number these days. What does it really mean? Does it mean anything at all? Is it just confusing us, trying to pull the wool over our eyes?

We can all learn and practise some simple rules of statistical thinking.

Statistics are provisional. Findings can always be improved. We must be willing to recognize the validity of the figures that are available today but ready to applaud rather than criticize when tomorrow there is better information available through which to revise them.

Statistics are uncertain. We should be sceptical of those peddling impossible guarantees, rather than demanding them, and celebrate those who tell us about risk and imprecision.

Statistics aim for impartiality. In a world of spin, where numbers are used to substantiate assertions rather than illuminate reality, we must be doggedly determined to root out bias and to ask ourselves whether something is really what it seems.

Statistics help us to join the dots. We must be modest in jumping to conclusions about cause and effect but appreciate the real value that comes from finding patterns that help us to ask good questions.

Everyone can do this. But not everyone can be a statistician. We need also to appreciate that those who are statisticians abide by principles of professionalism that should give confidence in what they say. We should look carefully at the provenance of statistical claims to see whether they are coming from people or organizations that have signed up to the necessary standards of conduct and have the necessary technical training and degree of expertise.

The principles of statistics have been captured in a variety of ways. The late Roger Jowell, who was a great mentor to me and an inspiration to many, was the leading light in the development of the International Statistical Institute's code of ethics (International Statistical Institute, 2010). Within this code is a set of ethical principles which set **a high bar for professional conduct:**

Principles

- (a) pursuing objectivity, without fear or favour and guarding against ‘predictable misinterpretation or misuse’;
- (b) clarifying obligations and roles, statisticians should take care to stay within their area of competence, and to seek advice, as appropriate, from others with the relevant expertise;
- (c) assessing alternatives impartially, assessing the respective merits and limitations of alternatives;
- (d) avoiding financial or personal conflicts of interest, exploring the likely consequences of collecting and disseminating various types of data and results;
- (e) rejecting any attempt to establish a predetermined outcome, including contractual conditions contingent on such an outcome;
- (f) guarding privileged information, although this prohibition is not to be extended to statistical methods and procedures utilized to conduct the inquiry or to produce published data;
- (g) exhibiting professional competence, upgrading professional knowledge and skills;
- (h) maintaining confidence in statistics, alerting potential users of the results to the limits of their reliability and applicability;
- (i) exposing and reviewing methods and findings through independent assessment;
- (j) communicating ethical principles;
- (k) bearing responsibility for the integrity of the discipline, not deceiving or knowingly misrepresenting or attempting to prevent reporting of misconduct or obstructing the scientific or scholarly research of others;
- (l) protecting the interests of subjects against potentially harmful effects of participation, including keeping confidential identities and records of all subjects or respondents.

THE FUTURE OF DATA ANALYSIS¹

By JOHN W. TUKEY

Princeton University and Bell Telephone Laboratories

I. General Considerations	2
1. Introduction	2
2. Special growth areas	3
3. How can new data analysis be initiated?	4
4. Sciences, mathematics, and the arts	5
5. Dangers of optimization	7
6. Why optimization?	8
7. The absence of judgment	9
8. The reflection of judgment upon theory	10
9. Teaching data analysis	11
10. Practicing data analysis	13
11. Facing uncertainty	13
II. Spotty Data	14
12. What is it?	14
13. An appropriate step forward	15
14. Trimming and Winsorizing samples	17
15. How soon should such techniques be put into service?	19
III. Spotty Data in More Complex Situations	21
16. Modified normal plotting	21
17. Automated examination	22
18. FUNOP	22
19. FUNOR-FUNOM in a two-way table	24
20. Example of use of FUNOR-FUNOM	27
IV. Multiple-Response Data	32
21. Where are we, and why?	32
22. The case of two samples	33
23. Factor analysis: the two parts	35
24. Factor analysis: regression	36
25. Factor analysis: the middle lines	36
26. Taxonomy; classification; incomplete data	38
V. Some Other Promising Areas	39
27. Stochastic-process data	39
28. Selection and screening problems	40
29. External, internal, and confounded estimates of error	41
30. The consequences of half-normal plotting	42
31. Heterogeneous data	42
32. Two samples with unequal variability	43
VI. Flexibility of Attack	44
33. Choice of modes of expression	44
34. Sizes, nomination, budgeting	45
35. A caveat about indications	46

Received July 1, 1961.

¹ Prepared in part in connection with research sponsored by the Army Research Office through Contract DA-36-034-ORD-2207 with Princeton University. Reproduction in whole or part is permitted for any purpose of the United States Government.

between these conventional levels. To be useful, a conclusion procedure need not be precise. As working data analysts we need to recognize that this is so.

We must face up to the fact that, in any experimental science, *our certainty about what will happen in a particular situation does not usually come from directly applicable experiments or theory*, but rather comes mainly through analogy between situations which are *not known* to behave similarly. Data analysis has, of necessity, to be an experimental science, and needs therefore to adopt the attitudes of experimental science. As a consequence our choices of analytical approach will usually be guided by what is known about simpler or similar situations, rather than by what is known about the situation at hand.

Finally, we need to give up the vain hope that data analysis can be founded upon a logico-deductive system like Euclidean plane geometry (or some form of the propositional calculus) and to face up to the fact that *data analysis is intrinsically an empirical science*. Some may feel let down by this, may feel that if data analysis cannot be a logico-deductive system, it inevitably falls to the state of a crass technology. With them I cannot agree. It will still be true that there will be aspects of data analysis well called technology, but there will also be the hallmarks of stimulating science: intellectual adventure, demanding calls upon insight, and a need to find out "how things really are" by investigation and the confrontation of insights with experience.

46. How might data analysis be taught? If we carry the point of view set forth here to its logical conclusion, we would teach data analysis in a very different way from any that I know to have been tried. We would teach it like biochemistry, with emphasis on what we have learned, with some class discussion of how such things were learned perhaps, but with relegation of all question of detailed methods to the "laboratory work". If we carried through the analogy to the end, all study of detailed proofs, as well as all trials of empirical sampling or comparisons of ways of presentation would belong in "the laboratory" rather than "in class". Moreover, practice in the use of data analysis techniques would be left to other courses in which problems arose, just as applications of biochemistry are left to other courses.

It seems likely, but not certain, that this would prove to be too great a switch to consider putting into immediate effect. Even if it is too much for one step, what about taking it in two or three steps?

I can hear the war cry "cookbookery" being raised against such a proposal. If raised it would fail, because the proposal is really to go in the opposite direction from cookbookery; to teach not "what to do", nor "how we learned what to do", but rather "what we have learned". This last is at the opposite pole from "cookbookery", goes beyond "the conduct of taste-testing panels", and is concerned with "the art of cookery". Dare we adventure?

47. The impact of the computer. How vital, and how important, to the matters we have discussed is the rise of the stored-program electronic computer? In many instances the answer may surprise many by being "important but not

In data analysis we find the “...hallmarks of stimulating science: intellectual adventure, demanding calls upon insight, and **a need to find out ‘how things really are’ by investigation and the confrontation of insights with experience.**”



REVISED AND UPDATED 4TH EDITION

BILL KOVACH
& TOM ROSENSTIEL

— THE —
ELEMENTS
OF
JOURNALISM

What Newspeople Should Know
and the Public Should Expect

In their book [The Elements of Journalism](#), Bill Kovach and Tom Rosenstiel identify the essential principles and practices of journalism.

Here are 10 elements common to good journalism, drawn from the book.

Journalism's first obligation is to the truth

Good decision-making depends on people having reliable, accurate facts put in a meaningful context. Journalism does not pursue truth in an absolute or philosophical sense, but in a capacity that is more down to earth.

"All truths – even the laws of science – are subject to revision, but we operate by them in the meantime because they are necessary and they work," Kovach and Rosenstiel write in the book. Journalism, they continue, thus seeks "a practical and functional form of truth." It is not the truth in the absolute or philosophical or scientific sense but rather a pursuit of "**the truths by which we can operate on a day-to-day basis.**"

This "journalistic truth" is a process that begins with the professional discipline of assembling and verifying facts. Then journalists try to convey a fair and reliable account of their meaning, subject to further investigation.

Journalists should be as transparent as possible about sources and methods so audiences can make their own assessment of the information. Even in a world of expanding voices, "getting it right" is the foundation upon which everything else is built – context, interpretation, comment, criticism, analysis and debate. The larger truth, over time, emerges from this forum.

As citizens encounter an ever-greater flow of data, they have more need – not less – for suppliers of information dedicated to finding and verifying the news and putting it in context.

Its first loyalty is to citizens

The publisher of journalism – whether a media corporation answering to advertisers and shareholders or a blogger with his own personal beliefs and priorities – must show an ultimate allegiance to citizens. **They must strive to put the public interest – and the truth – above their own self-interest or assumptions.**

A commitment to citizens is an implied covenant with the audience and a foundation of the journalistic business model – journalism provided “without fear or favor” is perceived to be more valuable than content from other information sources.

Commitment to citizens also means journalism should seek to present a representative picture of constituent groups in society. Ignoring certain citizens has the effect of disenfranchising them.

The theory underlying the modern news industry has been the belief that credibility builds a broad and loyal audience and that economic success follows in turn. **In that regard, the business people in a news organization also must nurture – not exploit – their allegiance to the audience ahead of other considerations.**

Technology may change but trust – when earned and nurtured – will endure.

Its essence is a discipline of verification

Journalists rely on a professional discipline for verifying information.

While there is no standardized code as such, every journalist uses certain methods to assess and test information to “get it right.”

Being impartial or neutral *is not* a core principle of journalism. **Because the journalist must make decisions, he or she is not and cannot be objective. But journalistic methods are objective.**

When the concept of objectivity originally evolved, it did not imply that journalists were free of bias. It called, rather, for **a consistent method of testing information – a transparent approach to evidence – precisely so that personal and cultural biases would not undermine the accuracy of the work.** The method is objective, not the journalist.

Seeking out multiple witnesses, disclosing as much as possible about sources, or asking various sides for comment, all signal such standards. This discipline of verification is what separates journalism from other forms of communication such as propaganda, advertising, fiction, or entertainment.

Its practitioners must maintain an independence from those they cover

Independence is a cornerstone of reliability.

On one level, it means not becoming seduced by sources, intimidated by power, or compromised by self-interest. **On a deeper level it speaks to an independence of spirit and an open-mindedness and intellectual curiosity that helps the journalist see beyond his or her own class or economic status, race, ethnicity, religion, gender or ego.**

Journalistic independence, write Kovach and Rosenstiel, is not neutrality. While editorialists and commentators are not neutral, the source of their credibility is still their accuracy, intellectual fairness and ability to inform – not their devotion to a certain group or outcome. In our independence, however, journalists must avoid straying into arrogance, elitism, isolation or nihilism.

It must serve as an independent monitor of power

Journalism has an unusual capacity to serve as watchdog over those whose power and position most affect citizens. **It may also offer voice to the voiceless.** Being an independent monitor of power means “watching over the powerful few in society on behalf of the many to guard against tyranny,” Kovach and Rosenstiel write.

The watchdog role is often misunderstood, even by journalists, to mean “afflict the comfortable.” While upsetting the applecart may certainly be a result of watchdog journalism, the concept as introduced in the mid-1600s was far less combative. Rather, it sought to redefine the role of the journalist from a passive stenographer to more **a curious observer who would “search out and discover the news.”**

The watchdog role also means more than simply monitoring government. “The earliest journalists,” write Kovach and Rosenstiel, “firmly established as a core principle their responsibility to examine unseen corners of society. The world they chronicled captured the imagination of a largely uninformed society, creating an immediate and enthusiastic popular following.”

Finally, the purpose of the watchdog extends beyond simply making the management and execution of power transparent, to making known and understood the effects of that power. This includes reporting on successes as well as failures.

Journalists have an obligation to protect this watchdog freedom by not demeaning it in frivolous use or exploiting it for commercial gain.

It must provide a forum for public criticism and compromise

The news media are common carriers of public discussion, and this responsibility forms a basis for special privileges that news and information providers receive from democratic societies.

These privileges can involve subsidies for distribution or research and development (lower postal rates for print, use of public spectrum by broadcasters, development and management of the Internet) to laws protecting content and free speech (copyright, libel, and shield laws).

These privileges, however, are not pre-ordained or perpetual. Rather, they are conferred because of the need for an abundant supply of information. They are predicated on the assumption that journalism – because of its principles and practices – will supply a steady stream of higher quality content that citizens and government will use to make better decisions.

Traditionally, this covenant has been between news organizations and government. The new forms of digital media, however, place a responsibility on everyone who “publishes” content – whether for profit or for personal satisfaction – in the public domain.

The raw material cast into the marketplace of ideas sustains civic dialogue and serves society best when it consists of verified information rather than just prejudice and supposition.

Journalism should also attempt to fairly represent varied viewpoints and interests in society and to place them in context rather than highlight only the conflicting fringes of debate. Accuracy and truthfulness also require that the public discussion not neglect points of common ground or instances where problems are not just identified but also solved.

Journalism, then, is more than providing an outlet for discussion or adding one’s voice to the conversation. **Journalism carries with it a responsibility to improve the quality of debate by providing verified information and intellectual rigor.** A forum without regard for facts fails to inform and degrades rather than improves the quality and effectiveness of citizen decision-making.

It must strive to keep the significant interesting and relevant

Journalism is storytelling with a purpose. It should do more than gather an audience or catalogue the important. It must balance what readers know they want with what they cannot anticipate but need.

Writing coaches Roy Peter Clark and Chip Scanlan describe effective newswriting as the intersection of civic clarity, the information citizens need to function, and literary grace, which is the reporter's storytelling skill set. In other words, part of the journalist's responsibility is providing information in such a way people will be inclined to listen. **Journalists must thus strive to make the significant interesting and relevant.**

Quality is measured both by how much a work engages its audience and enlightens it. This means journalists must continually ask what information has the most value to citizens and in what form people are most likely to assimilate it. While journalism should reach beyond such topics as government and public safety, journalism overwhelmed by trivia and false significance trivializes civic dialogue and ultimately public policy.

It must keep the news comprehensive and proportional

Journalism is our modern cartography. It creates a map for citizens to navigate society.

As with any map, its value depends on a completeness and proportionality in which the significant is given greater visibility than the trivial.

Keeping news in proportion is a cornerstone of truthfulness. Inflating events for sensation, neglecting others, stereotyping, or being disproportionately negative all make a less reliable map. **The most comprehensive maps include all affected communities, not just those with attractive demographics.** The most complete stories take into account diverse backgrounds and perspectives.

Though proportion and comprehensiveness are subjective, their ambiguity does not lessen their significance.

Its practitioners must be allowed to exercise their personal conscience

Doing journalism, whether as a professional writing for a news organization or as an online contributor in the public space, involves **one's moral compass and demands a personal sense of ethics and responsibility.**

Because “news” is important, those who provide news have a responsibility to voice their personal conscience out loud and allow others to do so as well. They must be willing to question their own work and to differ with the work of others if fairness and accuracy demand they do so.

News organizations do well to nurture this independence by encouraging individuals to speak their minds. Conversation and debate stimulate the intellectual diversity of minds and voices necessary to understand and accurately cover an increasingly diverse society. Having a diverse newsroom does little if those different voices are not spoken or heard.

It's also a matter of self-interest. Employees encouraged to raise their hands may “save the boss from himself” or protect the news organization's reputation by pointing out errors, flagging important omissions, questioning misguided assumptions, or even revealing wrongdoing.

Having a sense of ethics is perhaps most important for the individual journalist or online contributor.

Increasingly, those who produce “the news” work in isolation, whether from a newsroom cubicle, the scene of a story, or their home office. They may file directly to the public without the safety net of editing, a second set of eyes, or the collaboration of others. While crowdsourcing by the audience may catch and correct errors or misinformation, the reputation of the author and the quality of public dialogue are nevertheless damaged.

Citizens, who shape news production by the choices they make, have rights when it comes to news, but they also have responsibilities — even more so as they become producers and editors themselves.

The average person now, more than ever, works like a journalist.

Writing a blog entry, commenting on a social media site, sending a tweet, or “liking” a picture or post, likely involves a shorthand version of the journalistic process. One comes across information, decides whether or not it’s believable, assesses its strength and weaknesses, determines if it has value to others, decides what to ignore and what to pass on, chooses the best way to share it, and then hits the “send” button.

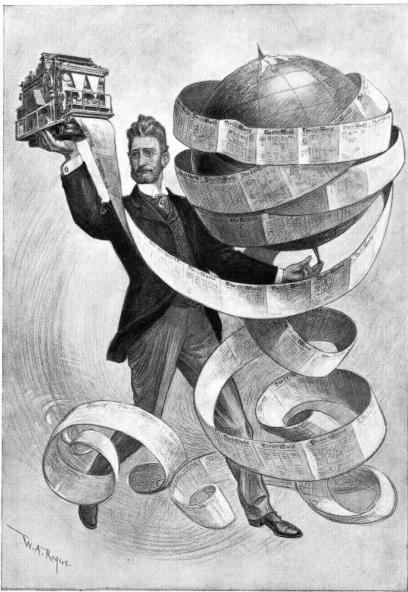
Though this process may take only a few moments, it’s essentially what reporters do.

Two things, however, separate this journalistic-like process from an end product that *is* “journalism.” The first is motive and intent. The purpose of journalism is to give people the information they need to make better decisions about their lives and society. The second difference is that journalism involves the conscious, systematic application of a discipline of verification to produce a “functional truth,” as opposed to something that is merely interesting or informative. Yet while the process is critical, it’s the end product – the “story” – by which journalism is ultimately judged.

Today, when the world is awash in information and news is available any time everywhere, a new relationship is being formed between the suppliers of journalism and the people who consume it.

The new journalist is no longer a gatekeeper who decides what the public should and should not know. The individual is now his or her own circulation manager and editor. To be relevant, journalists must now verify information the consumer already has or is likely to find and then help them make sense of what it means and how they might use it.

Thus, write Kovach and Rosenstiel, “**The first task of the new journalist/sense maker is to verify what information is reliable and then order it so people can grasp it efficiently.**” A part of this new journalistic responsibility is “to provide citizens with the tools they need to extract knowledge for themselves from the undifferentiated flood of rumor, propaganda, gossip, fact, assertion, and allegation the communications system now produces.”



AMERICAN EDITORS. II.—JOSEPH PULITZER

© 2001 HARPWEEK®

The use of data and computation in journalism is not new. Joseph Pulitzer himself wrote about how important "data" was to journalism.

NORTH AMERICAN REVIEW

No. DLXX.

MAY, 1904.

THE COLLEGE OF JOURNALISM.

A Review of Criticisms and Objections—Reflections Upon the Power, the Progress and the Prejudices of the Press—Why Specialized Concentration and Education at College Would Improve the Character and Work of Journalists and So Promote the Welfare of the Republic.

"The man who writes, the man who month in and month out, week in and week out, day in and day out, furnishes the material which is to shape the thoughts of our people, is essentially the man who more than any other determines the character of the people and the kind of government this people shall possess."

—PRESIDENT ROOSEVELT, April 7, 1904.

BY JOSEPH PULITZER.

THE editor of the NORTH AMERICAN REVIEW has asked me to reply to an article recently printed in its pages criticising the College of Journalism which it has been my pleasure to found and permanently to endow in Columbia University. In complying with his request I have enlarged the scope of the reply to include all other criticisms and misgivings, many honest, some shallow, some based on misunderstanding, but the most representing only prejudice and ignorance. If my comment upon these criticisms shall seem to be diffuse and perhaps repetitious, my apology is that—alas!—I am compelled to write by voice, not by pen, and to revise the proofs by ear, not by eye—a somewhat difficult task.

Some of my critics have called my scheme "visionary." If it

VOL. CLXXVIII.—NO. 570.

41

Copyright, 1904, by THE NORTH AMERICAN REVIEW PUBLISHING COMPANY. All Rights Reserved.

642

THE NORTH AMERICAN REVIEW.

be so I can at least plead that it is a vision I have cherished long, thought upon deeply and followed persistently. Twelve years ago I submitted the idea to President Low of Columbia, when it was declined. I have ever since continued to perfect and organize the scheme in my mind, until it is now accepted. In examining the criticisms and misgivings I have endeavored to do so without prejudice, anxious only to find the truth. I admit that the difficulties are many, but after weighing them all impartially I am more firmly convinced than ever of the ultimate success of the idea. Before the century closes schools of journalism will be generally accepted as a feature of specialized higher education, like schools of law or of medicine.

And now for our critics and objectors:

MUST A JOURNALIST BE "BORN?"

They object, the critics and cavillers, that a "newspaper man" must depend solely upon natural aptitude, or, in the common phrase, that he must be "born, not made."

Perhaps the critics can name some great editor, born full-winged like Mercury, the messenger of the gods? I know of none. The only position that occurs to me which a man in our Republic can successfully fill by the simple fact of birth is that of an idiot. Is there any other position for which a man does not demand and receive training—training at home, training in schools and colleges, training by master craftsmen, or training through bitter experience—through the burns that make the child dread the fire, through blunders costly to the aspirant?

This last is the process by which the profession of journalism at present obtains its recruits. It works by natural selection and the survival of the fittest, and its failures are strewn along the wayside.

The "born editor" who has succeeded greatly without special preparation is simply a man with unusual ability and aptitude for his chosen profession, with great power of concentration and sustained effort. He is one who loves his work and puts his whole heart and mind into it. He is in the strictest sense an educated man, but he has merely substituted self-education for education by others, making up for any deficiencies in his training by the unreserved sacrifice of strength, energy and pleasure. Even in his case might it not be an advantage to have a system of instruction that

In data he wrote we could find "romance, human interest, humor and fascinating revelations."

In ways that Pulitzer probably could not have imagined, over the last few decades, data and computing have become part of our everyday lives.

They regulate and shape our interactions with the physical and virtual worlds. Organizations increasingly equate (though not without problems)

"data release" with transparency. Sensing (sound, light, air quality) is cheap and plentiful, and easily deployed by the general public. Our actions online generate vast quantities of digital data.

And increasingly, computer systems exercise real power in the world through the insertion of machine learning (statistical models, artificial intelligence) alongside or in place of human decision making. In all of this, we can find new ways to ask questions about the world, how it's organized and how it functions. But the keys to this new digitized kingdom are data, code and algorithms. The curiosity, the questioning spirit, journalism students develop in their reporting classes finds an outlet in new and unexpected ways, mediated by data, code and algorithms.

Hence, a class in computational journalism. It is simply a response to our new condition of living in a computational society.

ciliation, of permanent State arbitration tribunals, of standing arbitration agreements between labor-unions and employers, and of the long line of international settlements leading up to the establishment of the world's court of arbitration at The Hague, would furnish material in themselves for a full and most valuable course of study for a journalist.

STATISTICS.

Everybody says that statistics should be taught. But how?

Statistics are not simply figures. It is said that nothing lies like figures—except facts. You want statistics to tell you the truth. You can find truth there if you know how to get at it, and romance, human interest, humor and fascinating revelations as well. The journalist must know how to find all these things—truth, of course, first. His figures must bear examination. It is much better to underestimate than to overstate his case, so that his critics and not himself may be put to confusion when they challenge him to verify his comparisons.

He must not read his statistics blindly; he must be able to test them by knowledge and by common sense. He must always be on the alert to discover how far they can actually be trusted—and what they really mean. The analysis of statistics to get at the essential truth of them has become a well-developed science, whose principles are systematically taught. And what a fascinating science it is! What romance can equal the facts of our national growth?

STATISTICIAN'S BRIEFING

EXERCISE: GUN SALES

Introducing some of the complexities of
Computational Journalism



Shootings never stopped during the pandemic: 2020 was the deadliest gun violence year in decades



A police officer stands near the scene of an afternoon shooting in Brooklyn that left one person dead in July. (Spencer Platt/Getty Images)

By Reis Thebault and Danielle Rindler

March 23, 2021 at 11:42 p.m. EDT

• ⌂ ⌁ ⌂ 1244

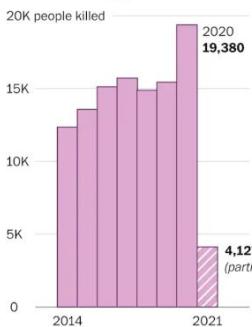
Until two lethal rampages this month, mass shootings had largely been absent from headlines during the [coronavirus](#) pandemic. But people were still dying — at a record rate.

In 2020, gun violence killed nearly 20,000 Americans, according to [data](#) from the Gun Violence Archive, more than any other year in [at least two decades](#). An additional 24,000 people died by suicide with a gun.

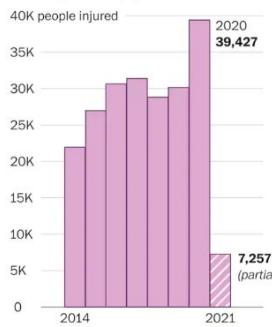
The vast majority of these tragedies happen far from the glare of the national spotlight, unfolding instead in homes or on city streets and — like the covid-19 crisis — disproportionately affecting communities of color.

2020 was a record year for gun violence

Gun deaths by year



Gun injuries by year



Note: Counts do not include the roughly 20,000-25,000 gun-related suicide deaths that occur each year.

Source: Gun Violence Archive

DANIELLE RINDLER/THE WASHINGTON POST

Last week's shootings at [spas in the Atlanta area](#) and Monday's shooting at a grocery store in Boulder, Colo., killed a combined 18 people and rejuvenated a national effort to overhaul gun laws. But high-profile mass shootings such as those tend to overshadow the instances of everyday violence that account for most gun deaths, potentially clouding some people's understanding of the problem and complicating the country's response, experts say.

"There are many communities across this country that are dealing with ever-present gun violence that is just part of their daily experience," said Mark Barden, a co-founder of the gun violence prevention group [Sandy Hook Promise](#). "It doesn't get the support, the spotlight, the national attention. People don't understand that it's continuous and it's on the rise."

Shooting deaths in 2020 outpaced the next-highest recent year, 2017, by more than 3,600. The rise resembles other alarming trends: Last year, the United States saw [the highest one-year increase](#) in homicides since it began keeping records, with [the country's largest cities](#) suffering a 30 percent spike. Gunshot injuries also rose dramatically, to nearly 40,000, over 8,000 more than in 2017.

"More than 100 Americans are killed daily by gun violence," Ronnie Dunn, a professor of urban studies at Cleveland State University, said, using a figure that includes suicides. "The majority are in Black and Brown communities. We don't really focus on gun violence until we have these mass shootings, but it's an ongoing, chronic problem that affects a significant portion of our society."



Police cars surround a Kum & Go gas station in Springfield, Mo., after a man killed four people before killing himself in March 2020. (Nathan Papes/Springfield News-Leader/AP)

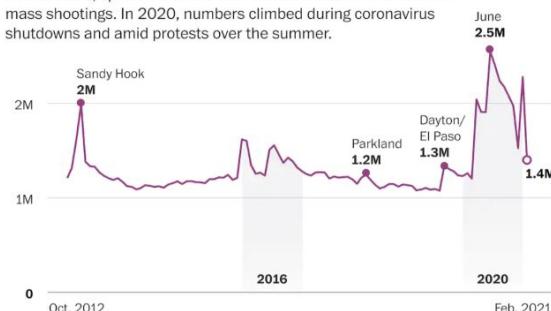
Researchers say the pandemic probably fueled the increases in several ways. The spread of the coronavirus hampered anti-crime efforts, and the attendant shutdowns compounded unemployment and stress at a time when schools and other

community programs were closed or online. They also note the apparent collapse of public confidence in law enforcement that followed [the police killing of George Floyd in Minneapolis](#).

Covid-19 and the protests over police brutality also led to [a surge of firearm sales](#). In 2020, people purchased about 23 million guns, a 64 percent increase over 2019 sales, according to a Washington Post analysis of federal data on gun background checks.

Estimated firearm sales surged in 2020

Before 2020, spikes coincided with elections and the months after mass shootings. In 2020, numbers climbed during coronavirus shutdowns and amid protests over the summer.



Note: Sales estimates are based on methodology surveying handgun, long gun, and multiple-gun background checks leading to purchases.

Source: Post analysis of FBI NICS data ANDREW BA TRAN AND DANIELLE RINDLER/THE WASHINGTON POST

Dunn pointed to this flood of firearms as the most detrimental factor in the fight to curb gun violence. When shootings become “the soundscape of inner-city neighborhoods,” he said, “it increases anxiety and stress and creates toxic stress.” Dunn compared the effect to post-traumatic stress disorder akin to what war veterans experience.

One recent study, from the Educational Fund to Stop Gun Violence, called gun violence “a public health crisis decades in the making.” An analysis of data from the Centers for Disease Control and Prevention found Black males between the ages of 15 and 34 accounted for 37 percent of gun homicides, even though they made up 2 percent of the U.S. population — a rate 20 times that of White males of the same age.

Failing to focus on this more common form of gun violence obscures the severity of the crisis, Dunn said.

Nicole Hockley is another co-founder of Sandy Hook Promise, who, like Barden, lost her first-grade son in a mass shooting at their elementary school in Newtown, Conn. She said she still recalls, with remorse, a time when she didn’t see the far-reaching impact of gun violence. When a gunman killed 12 people at a movie theater in Aurora, Colo., Hockley was ironing clothes in her living room.

“When I heard it on the news, my heart broke, I was so sad,” she said. “But then I got on with my life.”

Five months later, her son was killed at school.

“If we think about how many people are dying every single day, and we think about how would that be in our family or our community, perhaps that would motivate us to take ongoing action rather than spikes of action when something happens,” Hockley said.

“Even if it’s not happening in your community, it’s happening in the community of America.”



Shoppers are evacuated from a grocery store in Boulder, Colo., after a gunman opened fire on Monday. (Chet Strange/Getty Images)

Across the country, the swell of shootings has not spared the young.

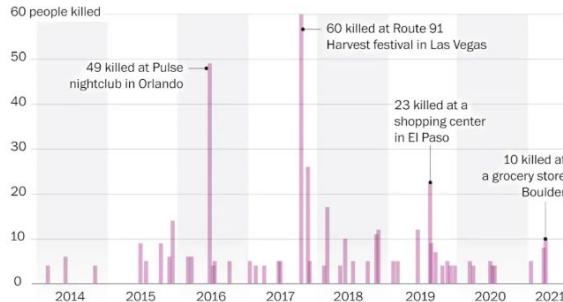
Nearly 300 children were shot and killed in 2020, according to Gun Violence Archive data, a 50 percent increase from the previous year. More than 5,100 kids and teens 17 and younger were killed or injured last year — over 1,000 more than any other year since 2014, when the website began tracking it.

The increase is especially striking because it occurred in a year when most children weren't attending class in person and were spared deadly school shootings. Experts say it highlights the severity of suicide and domestic violence.

[Sandy Hook Promise's crisis center](#) is fielding a record number of calls from young people contemplating suicide or witnessing other violence, Barden said.

"For a lot of students in our country, home is not the safest place to be," he said.

The rate of mass shootings has slowed during the pandemic, but they did not stop completely



40 people have died in 7 mass shootings since last March



Note: The Post defines a mass shooting as one that happens in a public place and in which four or more people, not including the shooter, are killed.

Source: Washington Post research, Gun Violence Archive DANIELLE RINDLER/THE WASHINGTON POST

undefined (Danielle Ridler)

Even though the rate of mass shootings slowed last year, several occurred before the killings in Atlanta and Boulder, according to The Post's [public mass shootings database](#).

Twenty-two people have been killed in five other shootings since last March: At a weekend Juneteenth celebration in Charlotte, a July 4 block party in Chicago and at a convenience store in Springfield, Mo., among others.

On average, there was one mass shooting every 73 days in 2020, compared with one every 36 days in 2019 and one every 45 days in 2017 and 2018. The slowdown interrupted what had been a five-year trend of more frequent and more deadly mass shootings.

That gun violence increased overall even as mass shootings declined underscores the fact that those high-profile events account for a relatively small share of firearm deaths. It should draw more attention to the victims and survivors of gun violence across the country, Barden said.

“They will bear the trauma and the scars of that catastrophe for the rest of their lives,” he said. “The collateral damage is unquantifiable, and it’s reaching almost everyone.”



Graham Bloomsmith and Cynthia Shi hold each other outside Gold Spa, where three women were shot and killed on March 16. (Elijah Nouvelage for The Washington Post)

BuzzFeedNews/nics-firearm-b... +

github.com/BuzzFeedNews/nics-firearm-background-checks

Search or jump to... Pull requests Issues Codespaces Marketplace Explore

BuzzFeedNews / nics-firearm-background-checks Public

Watch 25 Fork 123 Star 153

< Code Issues Pull requests 1 Actions Security Insights

master 3 branches 78 tags Go to file Add file Code About

jsvine Add December 2022 data and charts ab5a521 2 weeks ago 137 commits

charts	Add December 2022 data and charts	2 weeks ago
data	Add December 2022 data and charts	2 weeks ago
pdfs	Add December 2022 data and charts	2 weeks ago
scripts	Adjust script re. pandas FutureWarning	6 months ago
.gitignore	Add venv/ to .gitignore	last month
LICENSE.txt	Initial commit	7 years ago
Makefile	Update URL for NICS PDF.	6 years ago
README.md	Add December 2022 data and charts	2 weeks ago
requirements.txt	Update pinned versions in requirements.txt	6 months ago

README.md

FBI NICS Firearm Background Check Data

The data in this repository comes from the [FBI's National Instant Criminal Background Check System](#).

Mandated by the Brady Handgun Violence Prevention Act of 1993 and launched by the FBI on November 30, 1998, NICS is used by Federal Firearms Licensees (FFLs) to instantly determine whether a prospective buyer

About

Monthly data from the FBI's National Instant Criminal Background Check System, converted from PDF to CSV.

Readme MIT license 153 stars 25 watching 123 forks

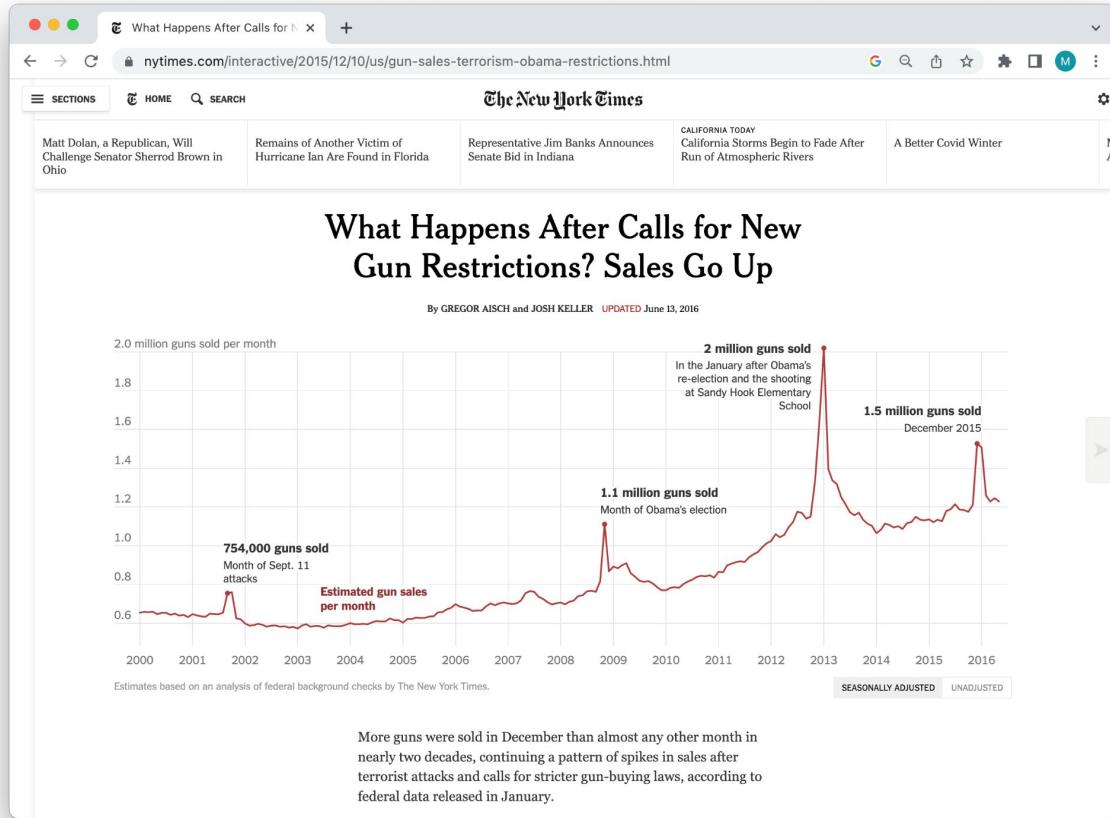
Releases 78 tags

Packages No packages published

Contributors 2

jsvine Jeremy Singer-Vine dependabot[bot]

<https://github.com/BuzzFeedNews/nics-firearm-background-checks>

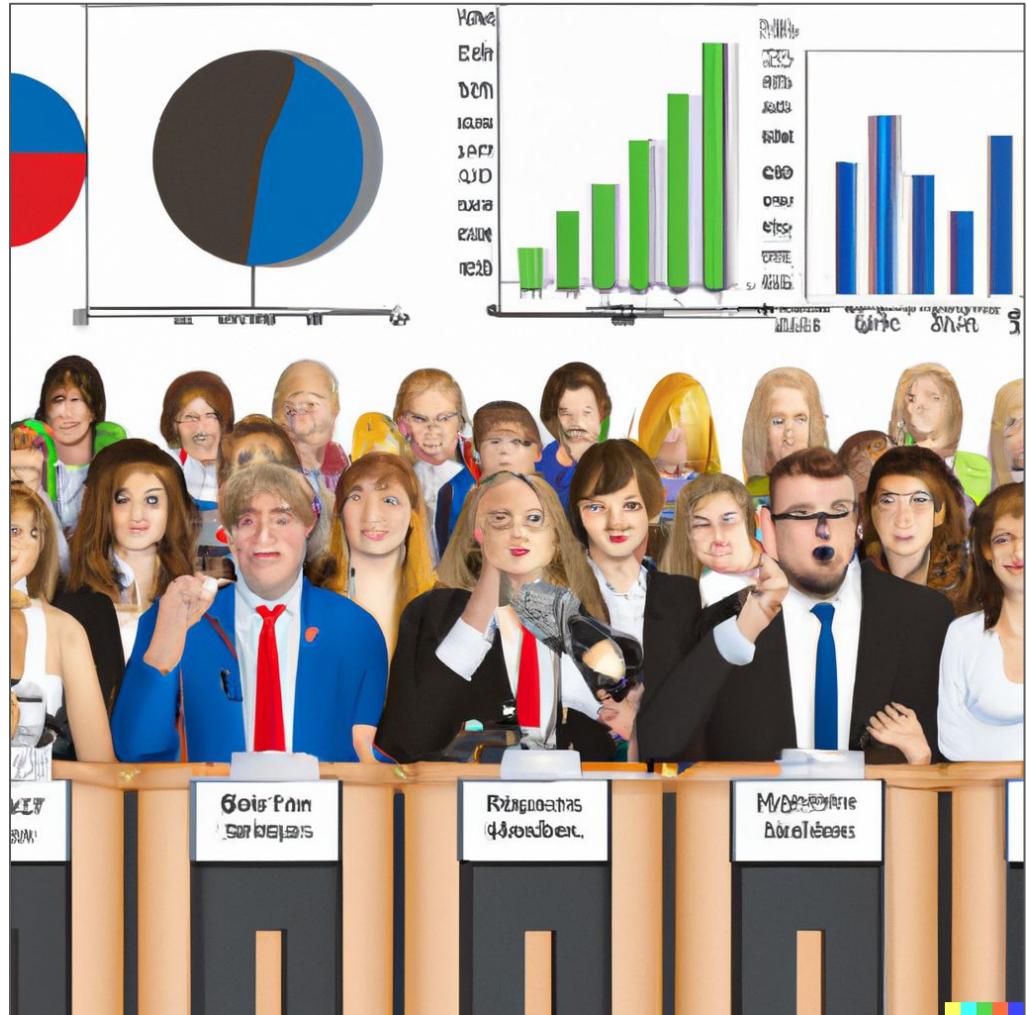


<https://www.nytimes.com/interactive/2015/12/10/us/gun-sales-terrorism-obama-restrictions.html>

EXERCISE: A LOOK AT CRIME STATISTICS

Introducing some of the complexities of Computational Journalism

STATISTICIAN'S BRIEFING



OCTOBER 31, 2022



Violent crime is a key midterm voting issue, but what does the data say?

BY JOHN GRAMLICH

Political candidates around the United States have released [thousands of ads](#) focusing on violent crime this year, and most registered voters see the issue as [very important](#) in the Nov. 8 midterm elections. But official statistics from the federal government paint a complicated picture when it comes to recent changes in the U.S. violent crime rate.

With Election Day approaching, here's a closer look at voter attitudes about violent crime, as well as an analysis of the nation's violent crime rate itself. All findings are drawn from Center surveys and the federal government's [two primary measures of crime](#): a large annual survey from the Bureau of Justice Statistics (BJS) and an annual study of local police data from the Federal Bureau of Investigation (FBI).

1 Around six-in-ten registered voters (61%) say violent crime is very important when making their decision about who to vote for in this year's congressional elections. Violent crime ranks alongside energy policy and health care in perceived importance as a midterm issue, but [far below the economy](#), according to the Center's October survey.

2 Republican voters are much more likely than Democratic voters to see violent crime as a key voting issue this year. Roughly three-quarters of Republican and GOP-leaning registered voters (73%) say violent crime is very important to their vote, compared with around half of Democratic or Democratic-leaning registered voters (49%).

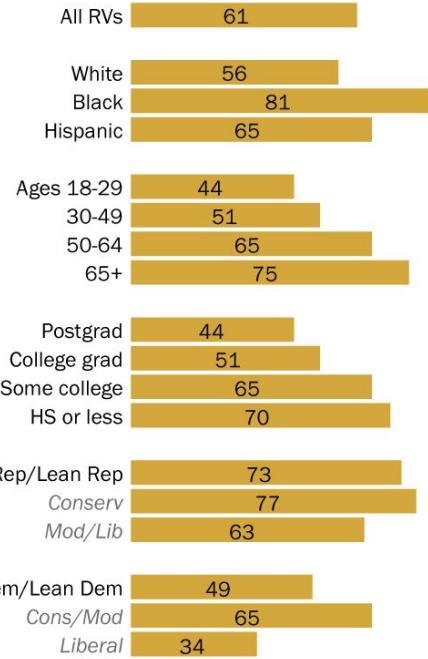
Conservative Republican voters are especially focused on the issue: About eight-in-ten (77%) see violent crime as very important to their vote, compared with 63% of moderate or liberal Republican voters, 65% of moderate or conservative Democratic voters and only about a third of liberal Democratic voters (34%).

3 Older voters are far more likely than younger ones to see violent crime as a key election issue. Three-quarters of registered voters ages 65 and older say violent crime is a very important voting issue for them this year, compared with fewer than half of voters under 30 (44%).

4 Black voters are particularly likely to say violent crime is a very important midterm issue. Black Americans have consistently been more likely than other racial and ethnic groups to express concern about violent crime, and that remains the case this year.

Some 81% of Black registered voters say violent crime is very important to their midterm vote, compared with 65% of Hispanic and 56% of White voters. (There were not enough Asian American voters in the Center's survey to analyze independently.)

Differences by race are especially pronounced among Democratic registered voters. While 82% of Black Democratic voters say violent crime is very important to their vote this year, only a third of White Democratic voters say the same.



Notes: Based on registered voters. There were not enough Asian American voters in the survey to analyze independently.
Source: Survey of U.S. adults conducted Oct. 10-16, 2022.

PEW RESEARCH CENTER

FBI Releases 2015 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2015-crime-statistics

An official website of the United States government. Here's how you know.

MORE ▾ ▶ > NEWS > PRESS RELEASES

FBI

NEWS

Stories | News Blog | Videos | Podcasts | **Press Releases** | Speeches | Testimony | Photos | Apps

Washington, D.C.

September 26, 2016

FBI Releases 2015 Crime Statistics

After two years of decline, the estimated number of violent crimes in the nation increased 3.9 percent in 2015 when compared with 2014 data, according to FBI figures released today. Property crimes dropped 2.6 percent, marking the 13th straight year the collective estimates for these offenses declined.

The 2015 statistics show the estimated rate of violent crime was 372.6 offenses per 100,000 inhabitants, and the property crime rate was 2,487.0 offenses per 100,000 inhabitants. The violent crime rate rose 3.1 percent compared with the 2014 rate, and the property crime rate declined 3.4 percent.

These and additional data are presented in the 2015 edition of the FBI's annual report *Crime in the United States*. This publication, which is a statistical compilation of offense, arrest, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's Uniform Crime Reporting (UCR) Program, also includes limited federal crime reporting, human trafficking, and cargo theft data.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and non-negligent manslaughter, rape, robbery, and aggravated assault, as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (Although the FBI classifies arson as a property crime, it does not estimate arson data because of variations in the level of participation by the reporting agencies. Consequently, arson data are not included in the property crime estimate.) The program also collects arrest data for the offenses listed above plus 20 offenses that include all other crimes except traffic violations.

Prior to 2013, the FBI's UCR Program collected rape data in the Summary Reporting System under the category "forcible rape." In 2013, the program removed the term "forcible" from the title and revised the definition. The legacy UCR definition of rape is "The carnal knowledge of a female forcibly and against her will." The revised UCR definition of rape is "Penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim."

Of the 18,439 city, county, university and college, state, tribal, and federal agencies eligible to participate in the UCR Program, 16,643 submitted data in 2015. A high-level summary of the statistics submitted, as well as estimates for those agencies that did not report, follows:

Twitter | Facebook | Email

In 2015, there were an estimated 1,197,704 violent crimes. Murder and non-negligent manslaughter increased 10.8 percent when compared with estimates from 2014.

Rape and aggravated assault increased 6.3 percent and 4.6 percent, respectively, while robbery increased 1.4 percent.

Nationwide, there were an estimated 7,993,631 property crimes. The estimated numbers for two of the three property crimes show declines when compared with the previous year's estimates. Burglaries dropped 7.8 percent, and larceny-thefts declined 1.8 percent, but motor vehicle thefts rose 3.1 percent.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$14.3 billion in 2015. The FBI estimated that law enforcement agencies nationwide made 10.8 million arrests, excluding traffic violations, in 2015.

The arrest rate for violent crime was 157.2 per 100,000 inhabitants, and the arrest rate for property crime was 458.9 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and non-negligent manslaughter was 3.5 per 100,000 inhabitants; rape (aggregate total of revised and legacy), 7.1; robbery, 29.7; and aggravated assault, 117.0 per 100,000 inhabitants.

The screenshot shows a web browser window with the title "Crime in Context | The Marshall Project". The main headline is "CRIME IN CONTEXT" in large, bold, black letters. Below it is a subtitle: "Violent crime is up in some places, but is it really a trend?". There are social media sharing icons for Facebook, Twitter, and Email. The author's name is listed as "By GABRIEL DANCE and TOM MEAGHER". The date and time of publication are "FILED 2:20 p.m. 08.18.2016". A note at the bottom states: "Updated 09.28.2016: This story has been updated with new data from the FBI's "Crime in the United States, 2015" report." The main text begins with a large, bold letter "I".

Is crime in America rising or falling? The answer is not nearly as simple as politicians sometimes make it out to be, because of how the FBI collects and handles crime data from the country's more than 18,000 police agencies. Those local reports are voluntary and sometimes inconsistent. And the bureau takes months or years to crunch the numbers, so the national data lags behind the current state of crime.

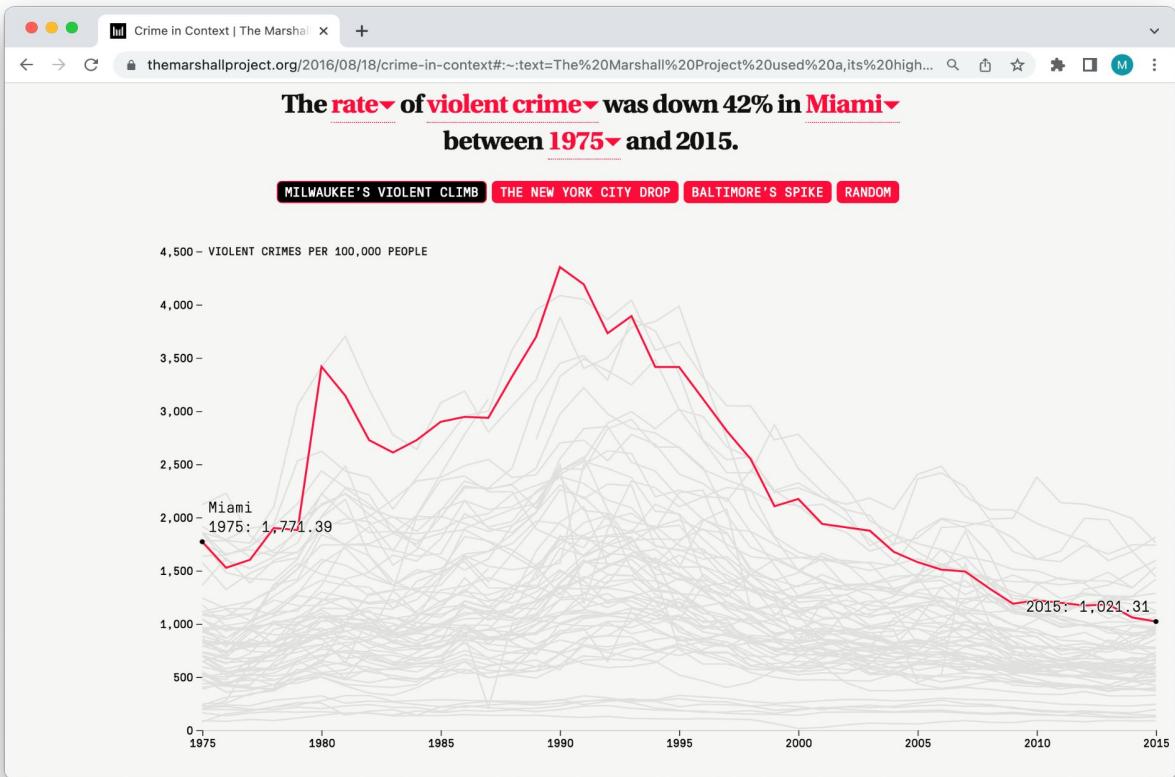
<https://www.themarshallproject.org/2016/08/18/crime-in-context>

A Matter of Perspective

In his speech accepting his party's nomination for president at the Republican National Convention in July, Donald Trump proclaimed that "Our president, who has used the pulpit of the presidency to divide us by race and color, has made America a more dangerous environment for everyone than frankly I have ever seen and anybody in this room has ever watched or seen."

But days later when the Democrats gathered for their convention in Philadelphia, President Obama responded, "Donald Trump calls [America] 'a divided crime scene' that only he can fix. It doesn't matter to him that illegal immigration and the crime rate are as low as they've been in decades..."

So who is right: Trump or Obama? Are we in the throes of a crime wave sweeping across the nation, or is this a period of stability and safety unlike any we've seen in a generation?



A Matter of Perspective

In his speech accepting his party's nomination for president at the Republican National Convention in July, Donald Trump proclaimed that "Our president, who has used the pulpit of the presidency to divide us by race and color, has made America a more dangerous environment for everyone than frankly I have ever seen and anybody in this room has ever watched or seen."

But days later when the Democrats gathered for their convention in Philadelphia, President Obama responded, "Donald Trump calls [America] 'a divided crime scene' that only he can fix. It doesn't matter to him that illegal immigration and the crime rate are as low as they've been in decades..."

So who is right: Trump or Obama? Are we in the throes of a crime wave sweeping across the nation, or is this a period of stability and safety unlike any we've seen in a generation?

FBI Releases 2016 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2016-crime-statistics

An official website of the United States government. Here's how you know.

MORE ▾ ▶ > NEWS > PRESS RELEASES

FBI

NEWS

Stories | News Blog | Videos | Podcasts | Press Releases | Speeches | Testimony | Photos | Apps

Washington, D.C.

September 25, 2017

FBI Releases 2016 Crime Statistics

The estimated number of violent crimes in the nation increased for the second straight year, rising 4.1 percent in 2016 when compared with 2015 data, according to FBI figures released today. Property crimes dropped 1.3 percent, marking the 14th consecutive year the collective estimates for these offenses declined.

The 2016 statistics show the estimated rate of violent crime was 386.3 offenses per 100,000 inhabitants, and the estimated rate of property crime was 2,450.7 offenses per 100,000 inhabitants. The violent crime rate rose 3.4 percent compared with the 2015 rate, and the property crime rate declined 2.0 percent.

These and additional data are presented in the 2016 edition of the FBI's annual report *Crime in the United States*. This publication is a statistical compilation of offense, arrest, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's Uniform Crime Reporting (UCR) Program. The UCR Program streamlined the 2016 edition by reducing the number of tables from 81 to 29, but still presented the major topics, such as offenses known, clearances, and persons arrested. Limited federal crime, human trafficking, and cargo theft data are also included.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and nonnegligent manslaughter, rape, robbery, and aggravated assault as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (Although the FBI classifies arson as a property crime, it does not estimate arson data because of variations in the level of participation by the reporting agencies. Consequently, arson data are not included in the property crime estimate.) The program also collects arrest data for the offenses listed above plus 20 offenses that include all other crimes except traffic violations.

Of the 18,481 city, county, university and college, state, tribal, and federal agencies eligible to participate in the UCR Program, 16,782 submitted data in 2016. A high-level summary of the statistics submitted, as well as estimates for those agencies that did not report, follows:

- In 2016, there were an estimated 1,248,185 violent crimes. Murder and nonnegligent manslaughter offenses increased 8.6 percent when compared with estimates from 2015. Aggravated assault and rape (legacy definition) offenses increased 5.1 percent and 4.9 percent,

In 2016, there were an estimated 1,248,185 violent crimes. Murder and nonnegligent manslaughter offenses increased 8.6 percent when compared with estimates from 2015. Aggravated assault and rape (legacy definition) offenses increased 5.1 percent and 4.9 percent, respectively, and robbery increased 1.2 percent.

Nationwide, there were an estimated 7,919,035 property crimes. The estimated numbers for two of the three property crimes show declines when compared with the previous year's estimates. Burglaries dropped 4.6 percent, larceny-thefts declined 1.5 percent, but motor vehicle thefts rose 7.4 percent.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$15.6 billion in 2016. The FBI estimated that law enforcement agencies nationwide made 10.7 million arrests, excluding those for traffic violations, in 2016.

The arrest rate for violent crime was 159.7 per 100,000 inhabitants, and the arrest rate for property crime was 420.6 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and nonnegligent manslaughter was 3.7 per 100,000 inhabitants; rape (aggregate total using the revised and legacy definition), 7.3; robbery, 29.8; and aggravated assault, 119.0 per 100,000 inhabitants.

FBI Releases 2017 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2017-crime-statistics

An official website of the United States government. Here's how you know.

MORE NEWS > PRESS RELEASES

NEWS

Stories | News Blog | Videos | Podcasts | Press Releases | Speeches | Testimony | Photos | Apps

Washington, D.C.

September 24, 2018

FBI Releases 2017 Crime Statistics

After two consecutive years of increases, the estimated number of violent crimes in the nation decreased 0.2 percent in 2017 when compared with 2016 data, according to FBI figures released today. Property crimes dropped 3.0 percent, marking the 15th consecutive year the collective estimates for these offenses declined.

The 2017 statistics show the estimated rate of violent crime was 382.9 offenses per 100,000 inhabitants, and the estimated rate of property crime was 2,362.2 offenses per 100,000 inhabitants. The violent crime rate fell 0.9 percent when compared with the 2016 rate; the property crime rate declined 3.6 percent.

These and additional data are presented in the 2017 edition of the FBI's annual report *Crime in the United States*. This publication is a statistical compilation of offense, arrest, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's Uniform Crime Reporting (UCR) Program.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and nonnegligent manslaughter, rape, robbery, and aggravated assault as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (The FBI classifies arson as a property crime, but it does not estimate arson data because of variations in the level of participation by the reporting agencies. Consequently, arson data is not included in the property crime estimate.) The program also collects arrest data for the offenses listed above plus 20 offenses that include all other crimes except traffic violations.

In 2013, the FBI's UCR Program initiated the collection of rape data under a revised definition within the Summary Reporting System. The term "forcible" was removed from the offense name, and the definition was changed to "penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim." In 2016, the FBI Director approved the recommendation to discontinue the reporting of rape data using the UCR legacy definition beginning in 2017.



In 2017, there were an estimated 1,247,321 violent crimes. The estimated number of robbery offenses decreased 4.0 percent, and the estimated number of murder and nonnegligent manslaughter offenses decreased 0.7 percent when compared with estimates from 2016. The estimated volume of aggravated assault and rape (revised definition) offenses increased 1.0 percent and 2.5 percent, respectively.

Nationwide, there were an estimated 7,694,086 property crimes. The estimated numbers for two of the three property crimes showed declines when compared with the previous year's estimates. Burglaries dropped 7.6 percent, larceny-thefts decreased 2.2 percent, but motor vehicle thefts rose 0.8 percent.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$15.3 billion in 2017. The FBI estimated law enforcement agencies nationwide made 10.6 million arrests, (excluding those for traffic violations) in 2017.

The arrest rate for violent crime was 160.7 per 100,000 inhabitants; the arrest rate for property crime was 388.7 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and nonnegligent manslaughter was 3.8 per 100,000 inhabitants; rape (aggregate total using the revised and legacy definition), 7.2; robbery, 29.3; and aggravated assault, 120.4 per 100,000 inhabitants.

FBI Releases 2018 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2018-crime-statistics

An official website of the United States government. Here's how you know.

MORE ▾ ▶ > NEWS > PRESS RELEASES

FBI

NEWS

Stories | News Blog | Videos | Podcasts | Press Releases | Speeches | Testimony | Photos | Apps

Washington, D.C.

September 30, 2019

FBI Releases 2018 Crime Statistics

For the second consecutive year, the estimated number of violent crimes in the nation decreased when compared with the previous year's statistics, according to FBI figures released today. In 2018, violent crime was down 3.3 percent from the 2017 number. Property crimes also dropped 6.3 percent, marking the 16th consecutive year the collective estimates for these offenses declined.

The 2018 statistics show the estimated rate of violent crime was 368.9 offenses per 100,000 inhabitants, and the estimated rate of property crime was 2,199.5 offenses per 100,000 inhabitants. The violent crime rate fell 3.9 percent when compared with the 2017 rate; the property crime rate declined 6.9 percent.

These and additional data are presented in the 2018 edition of the FBI's annual report *Crime in the United States*. This publication is a statistical compilation of offense, arrest, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's Uniform Crime Reporting (UCR) Program.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and nonnegligent manslaughter, rape, robbery, and aggravated assault, as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (The FBI classifies arson as a property crime, but does not estimate arson data because of variations in the level of participation by the reporting agencies. Consequently, arson data is not included in the property crime estimate.) The program also collects arrest data for the offenses listed above plus 20 offenses that include all other crimes except traffic violations.

Of the 18,586 federal, state, county, city, university and college, and tribal agencies eligible to participate in the UCR Program, 16,659 agencies submitted data in 2018. A high-level summary of the statistics submitted, as well as estimates for those agencies that did not report, follows:

- In 2018, there were an estimated 1,206,836 violent crimes. The estimated number of three violent crime offenses decreased when compared with estimates from 2017. Robbery offenses fell 12.0 percent, murder and nonnegligent manslaughter offenses fell 6.2 percent, and the estimated volume of aggravated assault offenses decreased 0.4 percent. The estimated volume of rape (revised

In 2018, there were an estimated 1,206,836 violent crimes. The estimated number of three violent crime offenses decreased when compared with estimates from 2017. Robbery offenses fell 12.0 percent, murder and nonnegligent manslaughter offenses fell 6.2 percent, and the estimated volume of aggravated assault offenses decreased 0.4 percent. The estimated volume of rape (revised definition) offenses increased 2.7 percent.

Nationwide, there were an estimated 7,196,045 property crimes. The estimated numbers for all three property crimes showed declines when compared with the previous year's estimates. Burglaries dropped 11.9 percent, larceny-thefts decreased 5.4 percent, and motor vehicle thefts were down 3.1 percent.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$16.4 billion in 2018. The FBI estimated law enforcement agencies nationwide made 10.3 million arrests, (excluding those for traffic violations) in 2018.

The arrest rate for violent crime was 159.9 per 100,000 inhabitants, and the arrest rate for property crime was 361.2 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and nonnegligent manslaughter was 3.7 per 100,000 inhabitants; rape (aggregate total using the revised and legacy definition), 7.7; robbery, 27.2; and aggravated assault, 121.4 per 100,000 inhabitants.

FBI Releases 2019 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2019-crime-statistics

An official website of the United States government. Here's how you know.

MORE NEWS > PRESS RELEASES

NEWS

Stories | News Blog | Videos | Podcasts | Press Releases | Speeches | Testimony | Photos | Apps

Washington, D.C.
FBI National Press Office
(202) 324-3691

September 28, 2020

FBI Releases 2019 Crime Statistics

For the third consecutive year, the estimated number of violent crimes in the nation decreased when compared with the previous year's statistics, according to FBI figures released today. In 2019, violent crime was down 0.5% from the 2018 number. Property crimes also dropped 4.1%, marking the 17th consecutive year the collective estimates for these offenses declined.

The 2019 statistics show the estimated rate of violent crime was 366.7 offenses per 100,000 inhabitants, and the estimated rate of property crime was 2,109.9 offenses per 100,000 inhabitants. The violent crime rate fell 1.0% when compared with the 2018 rate; the property crime rate declined 4.5%.

These and additional data are presented in the 2019 edition of the FBI's annual report *Crime in the United States*. This publication is a statistical compilation of offense, arrest, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's Uniform Crime Reporting (UCR) Program.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and nonnegligent manslaughter, rape, robbery, and aggravated assault, as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (The FBI classifies arson as a property crime but does not estimate arson data because of variations in the level of participation by the reporting agencies. Consequently, arson data is not included in the property crime estimate.) The program also collects arrest data for the offenses listed above and 20 offenses that include all other crimes except traffic violations.

Of the 18,667 federal, state, county, city, university and college, and tribal agencies eligible to participate in the UCR Program, 16,554 agencies submitted data in 2019. A high-level summary of the statistics submitted, as well as estimates for those agencies that did not report, follows:

- In 2019, there were an estimated 1,203,808 violent crimes. When compared with the estimates from 2018, the estimated number of

In 2019, there were an estimated 1,203,808 violent crimes. When compared with the estimates from 2018, the estimated number of robbery offenses fell 4.7% and the estimated volume of rape (revised definition) offenses decreased 2.7%. The estimated number of aggravated assault offenses rose 1.3%, and the volume of murder and nonnegligent manslaughter offenses increased 0.3%.

Nationwide, there were an estimated 6,925,677 property crimes. The estimated numbers for all three property crimes showed declines when compared with the previous year's estimates. Burglaries dropped 9.5%, larceny-thefts decreased 2.8%, and motor vehicle thefts were down 4.0%.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$15.8 billion in 2019.

The FBI estimated law enforcement agencies nationwide made 10.1 million arrests, (excluding those for traffic violations) in 2019.

The arrest rate for violent crime was 156.3 per 100,000 inhabitants, and the arrest rate for property crime was 343.3 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and nonnegligent manslaughter was 3.4 per 100,000 inhabitants; rape (aggregate total using the revised and legacy definition), 7.4; robbery, 24.7; and aggravated assault, 120.8 per 100,000 inhabitants.

FBI Releases 2020 Crime Stats

fbi.gov/news/press-releases/fbi-releases-2020-crime-statistics

An official website of the United States government. Here's how you know.

MORE ▾ > NEWS > PRESS RELEASES

NEWS

Stories News Blog Videos Podcasts Press Releases Speeches Testimony Photos Apps

Washington, D.C.
FBI National Press Office
(202) 324-3691

September 27, 2021

FBI Releases 2020 Crime Statistics

For the first time in four years, the estimated number of violent crimes in the nation increased when compared with the previous year's statistics, according to FBI figures released today. In 2020, violent crime was up 5.6 percent from the 2019 number. Property crimes dropped 7.8 percent, marking the 18th consecutive year the collective estimates for these offenses declined.

The 2020 statistics show the estimated rate of violent crime was 387.8 offenses per 100,000 inhabitants, and the estimated rate of property crime was 1,958.2 offenses per 100,000 inhabitants. The violent crime rate rose 5.2 percent when compared with the 2019 rate; the property crime rate declined 8.1 percent.

These additional data are presented in the 2020 edition of the FBI's annual report *Crime in the United States*. This report is available as downloadable spreadsheets and topic pages about offenses, arrests, and police employee data reported by law enforcement agencies voluntarily participating in the FBI's *Uniform Crime Reporting (UCR) Program*.

The UCR Program collects information on crimes reported by law enforcement agencies regarding the violent crimes of murder and nonnegligent manslaughter, rape, robbery, and aggravated assault, as well as the property crimes of burglary, larceny-theft, motor vehicle theft, and arson. (The FBI classifies arson as a property crime but does not estimate arson data because of variations in the level of

Crime in 2020

VIOLENT CRIMES
1,277,696
+5.6% from 2019

PROPERTY CRIMES
6,452,038
-7.8% from 2019

From *Crime in the United States, 2020*

In 2020, there were an estimated 1,277,696 violent crimes. When compared with the estimates from 2019, the estimated number of robbery offenses fell 9.3 percent and the estimated volume of rape (revised definition) offenses decreased 12.0 percent. The estimated number of aggravated assault offenses rose 12.1 percent, and the volume of murder and nonnegligent manslaughter offenses increased 29.4 percent.

Nationwide, there were an estimated 6,452,038 property crimes. The estimated numbers for two of the three property crimes showed declines when compared with the previous year's estimates. Burglaries dropped 7.4 percent, larceny-thefts decreased 10.6 percent, while motor vehicle thefts rose 11.8 percent.

Collectively, victims of property crimes (excluding arson) suffered losses estimated at \$17.5 billion in 2020. The FBI estimated law enforcement agencies nationwide made 7.6 million arrests, (excluding those for traffic violations) in 2020.

The arrest rate for violent crime was 147.9 per 100,000 inhabitants, and the arrest rate for property crime was 267.3 per 100,000 inhabitants.

By violent crime offense, the arrest rate for murder and nonnegligent manslaughter was 3.8 per 100,000 inhabitants; rape (aggregate total using the revised and legacy definition), 6.3; robbery, 21.0; and aggravated assault, 116.8 per 100,000 inhabitants.

[fbi.gov/news/press-releases/press-releases/fbi-releases-2021-crime-in-the-nation-statistics](https://www.fbi.gov/news/press-releases/press-releases/fbi-releases-2021-crime-in-the-nation-statistics)

MORE ▾ A > NEWS > PRESS RELEASES > PRESS RELEASES

FBI

NEWS

Stories | News Blog | Videos | Podcasts | **Press Releases** | Speeches | Testimony | Photos | Apps

Washington, D.C.
FBI National Press Office
(202) 324-3691

October 5, 2022

FBI Releases 2021 Crime in the Nation Statistics

Today, the FBI released detailed data on over 11 million criminal offenses reported to the Uniform Crime Reporting (UCR) Program's National Incident-Based Reporting System (NIBRS) in 2021 via *NIBRS, 2021; NIBRS Estimates, 2021; The Transition to the National Incident-Based Reporting System (NIBRS): A Comparison of 2020 and 2021 NIBRS Estimates; and Crime in the United States (CIUS), 2021*.

For years, the FBI's UCR Program has provided annual snapshots of crime in the nation. This year, users will notice a difference in the data because it was exclusively collected via NIBRS in 2021. Both the *NIBRS, 2021* and *CIUS, 2021* releases are based solely on these NIBRS submissions.

Establishing NIBRS as the national standard for crime data provides the opportunity to know more about, and better understand, various facets of crime in our nation. NIBRS provides an avenue for the UCR Program to estimate the amount of arson committed each year, estimates on drug offenses by drug type, and victimization estimates. NIBRS also provides estimates on victim and arrestee demographics, including age, sex, and race.

In anticipation of UCR's evolution to NIBRS, the FBI collaborated with the [Bureau of Justice Statistics \(BJS\)](#) to develop comprehensive methodologies to bring a NIBRS estimation process to fruition and establish 2021 as the first year in which all crime estimates can be based



Estimates Will Help Fill in Crime Statistics Gap

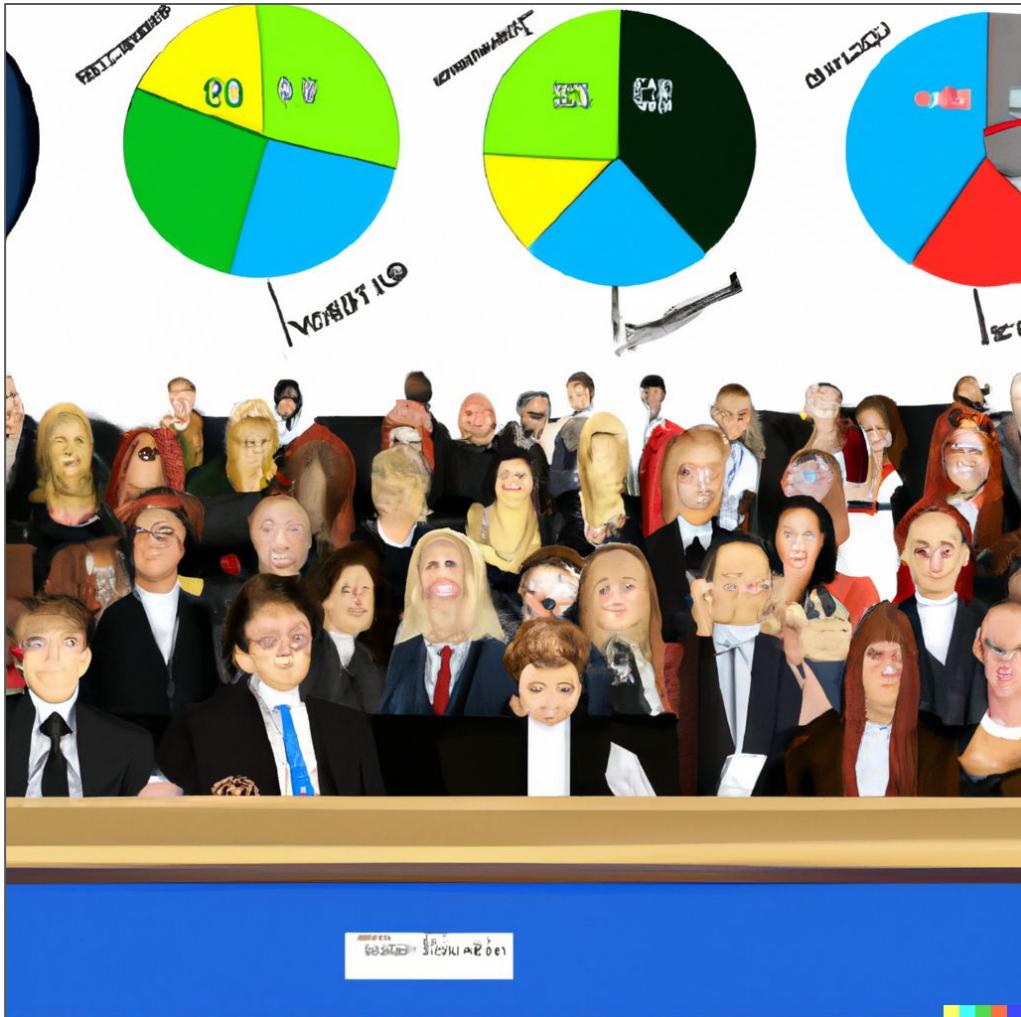
When the annual Crime in the United States data is released this fall, it will look different from previous releases.

What's the story?

EXERCISE: A LOOK AT CRIME STATISTICS

Introducing some of the complexities of
Computational Journalism

SPECIAL GUEST!



What Can FBI Data Say About Crime in 2021? It's Too Unreliable to Tell

The transition to a new data system creates huge gaps in national crime stats sure to be exploited by politicians in this election year.

By WEIHUA LI

[f](#) [t](#) [e](#) [p](#)

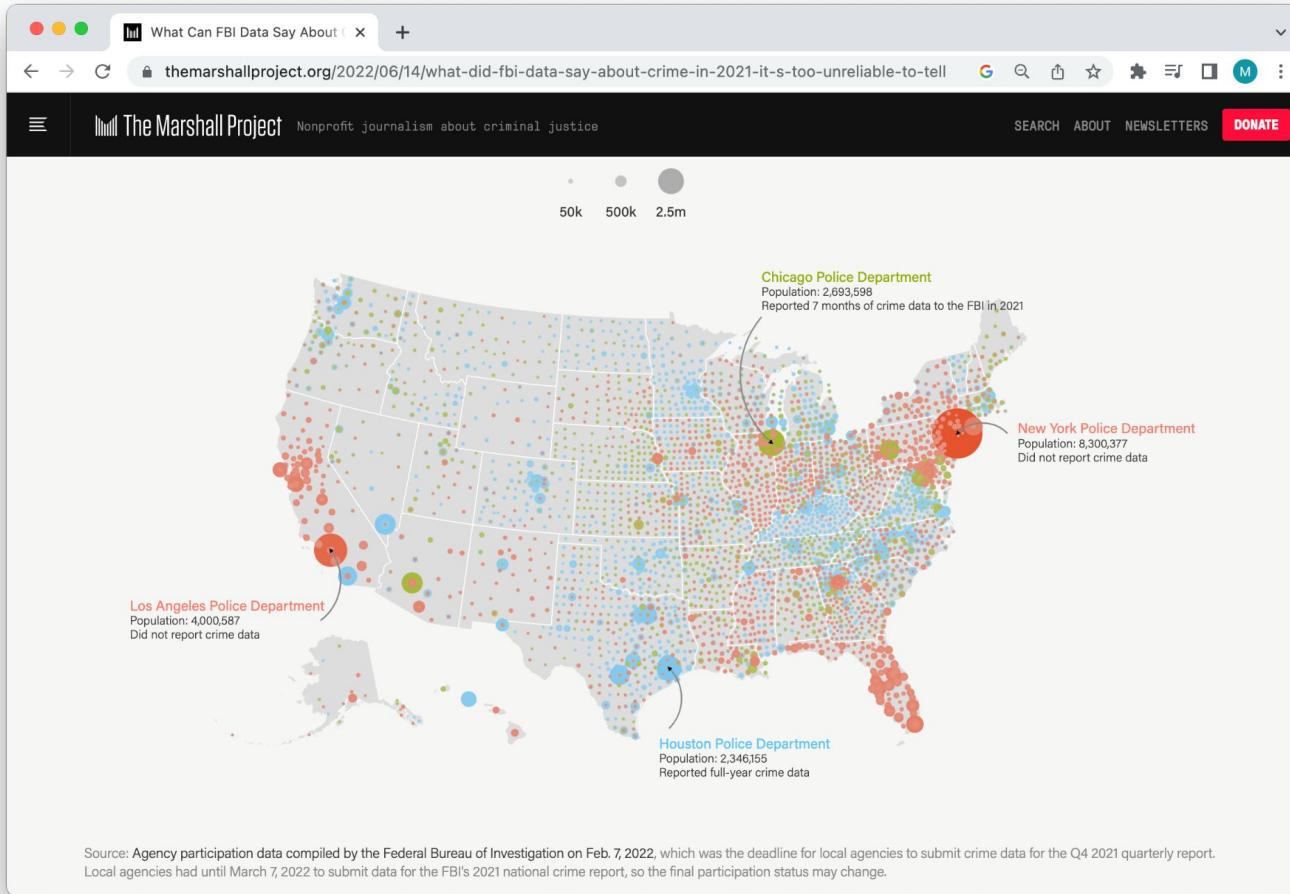
Nearly 40% of law enforcement agencies around the country did not submit any data in 2021 to a newly revised FBI crime statistics collection program, leaving a massive gap in information sure to be exploited by politicians in midterm election campaigns already dominated by public fear over a rise in violent crime.

This article was published in partnership with [Axios Local](#).

Download the agency participation data used in this piece.

The gap includes the nation's two largest cities by population, New York City and Los Angeles, as well as most agencies in five of the six most populous states: California, New York, Illinois, Pennsylvania, and Florida.

In 2021, the FBI retired its nearly century-old national crime data collection program, the Summary Reporting System used by the Uniform Crime Reporting (UCR) program. The agency switched to a new system, the National Incident-Based Reporting System (NIBRS), which gathers more specific information on each incident. Even though the FBI announced the transition years ago and the federal government spent hundreds of millions of dollars to help local police make the switch, about 7,000 of the nation's 18,000 law enforcement agencies did not successfully send crime data to the voluntary program last year.





Estimation Procedures for Crimes in the United States Based on NIBRS Data

AUGUST 2022

Marcus Berzofsky, Dan Liao, and G. Lance Couzens (RTI International)
Erica L. Smith (BJS)
Cynthia Barnett-Ryan (FBI-CJIS)

This report describes the estimation process for crime statistics that will be published by the FBI based solely on data submitted by state and local law enforcement agencies to the National Incident-Based Reporting System (NIBRS). It provides a basic background on general estimation and the use of confidence intervals around the estimates. *Estimation* refers to the statistical process that allows inferences to be made about an outcome of interest (e.g., the number of murders in the United States) in a population, even if information about the outcome is only known for part of the population.

Additionally, the report describes why estimation is needed to generate national statistics on reported crime based on NIBRS data, discusses how the new estimation process will differ from the previous process, and details how the new procedures account for the challenges faced when estimating crime based solely on NIBRS. This information will enable the public to interpret the new crime estimates and to understand that the estimates will continue to serve as a reliable and accurate source of information about crime known to law enforcement, regardless of changes to the statistical methodology or the underlying data source.

Introduction

Since 1930, the FBI has gathered and published annual crime statistics based on data voluntarily submitted by law enforcement to the Summary Reporting System (SRS) of the FBI's Uniform Crime Reporting (UCR) Program, providing an authoritative perspective on the scope of crime reported to law enforcement in the nation. The SRS data collection was voluntary and not all law enforcement agencies provided data each year. To account for agencies that did not submit data, the FBI began estimating crime in the 1960s, using the reports of participating agencies to produce national

and state crime estimates. The aggregate crime counts and estimates from the SRS served data users well over the years, but the growing need for more detailed information on crime known to law enforcement led to the development of NIBRS in the mid-1980s. After NIBRS was established, state crime reporting programs and local agencies could decide if they would report data using SRS or NIBRS. To accommodate that choice, the FBI's UCR Program collected crime and arrest data through both SRS and NIBRS, and annual national estimates of reported crime were based on the aggregation of both sources of data.

In 2016, with support from prominent law enforcement organizations, the FBI announced that the UCR Program would retire the SRS on January 1, 2021. Following that decision, the UCR Program modernized its data collection system and increased the frequency of crime data releases. The FBI, working closely with the Bureau of Justice Statistics (BJS), also encouraged its law enforcement partners to transition to NIBRS, with BJS and the FBI providing support through training and grant funding.

The FBI moved to a NIBRS-only data collection on January 1, 2021. The transition to NIBRS offers several significant improvements compared to the SRS:

1. *Counting all offenses in an incident*—When crime incident information is collected, each offense in an incident is counted. Comparatively, SRS imposed the hierarchy rule, where only the most serious offense in an incident was counted.¹

¹The hierarchy rule in the SRS states when more than one offense occurs within an incident, only the most serious crime contributes to the agency's monthly crime totals. For example, an incident involving murder, robbery, and motor vehicle theft only counts the homicide for the monthly totals, as homicide is the highest offense on the hierarchy. <https://ucr.fbi.gov/nibrs/2012/resources/effects-of-nibrs-on-crime-statistics>

Estimation Procedures for Crimes in the United States Based on NIBRS Data

2. *Providing a more complete picture of crime known to law enforcement*—NIBRS enables law enforcement agencies to report information on 58 different offense types, covering a wider variety of crimes than the 10 offenses formerly recorded in the SRS. NIBRS data provide a much more in-depth profile of reported crime occurring in a community.

3. *Capturing information about the characteristics of each crime incident*—NIBRS vastly expands the information collected on reported crime, to include demographic characteristics of crime victims, offenders, and persons arrested; details of the incident type and location; the types of weapons used; and important relationships between victims and offenders. SRS did not collect all of these specific details.

The transition to NIBRS means that the FBI's UCR Program will generate crime statistics based solely on data submitted to NIBRS. In 2021, approximately 66 percent of law enforcement agencies in the United States submitted data to NIBRS. To calculate a national crime rate based on NIBRS data, the FBI, in partnership with BJS and with statistical and technical support from RTI International, established a new set of statistical procedures to estimate crime using data provided by NIBRS-contributing agencies. Relying solely on NIBRS data for this new estimation process required addressing several challenges:

1. *Producing national and state-level crime estimates as law enforcement agencies continue to transition to NIBRS*—The number of agencies transitioning to NIBRS has increased annually since 2016, when the FBI announced their intention to retire the SRS as of January 1, 2021. In addition, in 2021, the first year for which estimates will be produced, the overall number of law enforcement agencies submitting data to NIBRS is smaller than the number of agencies for which SRS data were available in prior years. The estimation methodology must account for the reduced number of agencies reporting in 2021, as well as be flexible to accommodate the fluctuation in the number of reporting agencies in future years.

2. *Developing estimation techniques that can run effectively and efficiently on very large amounts of data*—The NIBRS data structure is much larger than SRS, as each NIBRS record contains a rich set of information on each crime incident, the victims affected, and the offenders and arrested persons involved. The SRS data generally consisted of

aggregate offense counts with limited supplemental information for each law enforcement agency. Conversely, NIBRS reporting agencies provide incident-specific records for all the recorded crime incidents in their jurisdiction. The ability to connect the details within each incident—for instance, victim data with the criminal offenses they experienced—provides context about crime that was previously unavailable. This means that one agency could potentially report thousands of individual crime incidents through NIBRS, whereas they would have reported only summarized offense totals to SRS. The increased size and detail of the NIBRS database requires, therefore, a more complex estimation process.

3. *Identifying the set of key indicators for which estimates will be produced*—The expanded level of detail in NIBRS could hypothetically result in the production of several million more estimates compared to SRS. The new estimation methodology needs to target a set of priority indicators of crime and arrest to ensure the procedures generate valid and reliable estimates on these key metrics based on complete and high-quality data.

Estimation Basics

Estimation enables the conversion of statistical sample data into estimates of population characteristics. It is the statistical process which allows inferences to be made about an outcome of interest in a population (e.g., the number of murders in the United States), even if information about the outcome is only known for part of the population. Regarding data submitted to the UCR Program, *inferences* are made about crimes known to law enforcement in the United States. Not all law enforcement agencies submit all their crime data; some provide only partial data, or no data at all, on crimes and arrests. Estimation is used to generate statistics about crime known to law enforcement that are still representative of the entire population, including the population covered by the agencies that did not provide complete information.

Estimation is a statistical process based on data from a subset of the population, which means that any estimated statistic has some amount of uncertainty associated with it. Uncertainty is a combination of the natural, random variation in an indicator plus any variation resulting from specific, identifiable, or systematic causes that over- or under-measure that indicator. Both types of variation are often used to

The Counted: people killed by police in the US

The Counted

People killed by police in the US

SEND A TIP DATABASE ABOUT READ ARTICLES JOIN US: [f](#) [t](#) [e](#)

PEOPLE KILLED IN 2016

RACE & ETHNICITY PER MILLION TOTAL

Native American	10.13	10.13
Black	6.66	6.66
Hispanic/Latino	3.23	3.23
White	2.9	2.9
Asian/Pacific Islander	1.17	1.17

STATE NAME PER CAPITA TOTAL

AK	NM	OK	DC	AR	AZ	WV	SD	CO	AL	ALABAMA	Total killed:	27
NV	KY	HI	MT	LA	OR	NE	CA	TN	MO		Population:	4,858,979
KS	WA	SC	MS	FL	NC	WY	TX	VT	WI		Rank (per capita):	10 / 51
IN	ID	GA	UT	MD	MN	OH	VA	IL	MA		Rank (total):	11 / 51
RI	PA	MI	IA	NJ	ME	NH	CT	ND	NY			
DE												

f t

2016 2015 List Map

Search by name:

e.g Freddie Gray

Filter by:

State

Gender

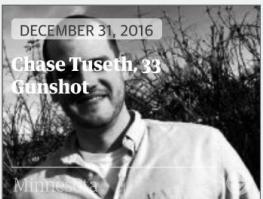
Race/ethnicity

Armed/unarmed

Age

Classification

December 2016



85 people

washingtonpost.com/graphics/investigations/police-shootings-database/

The Washington Post
Democracy Dies in Darkness

markh

FATAL FORCE

1,101 people have been shot and killed by police in the past 12 months

The Post has tracked 8,124 fatal police shootings since 2015

Donna Dale, 47
killed on Jan. 11, 2023
in Florida

Jackie L. Haynes, 52
killed on Jan. 9, 2023
in Newnan, Ga.

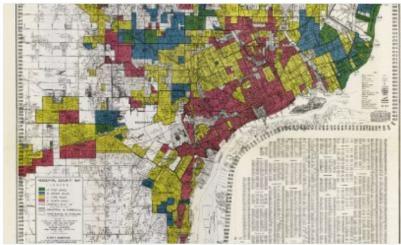
Chiewelthap Mariar, 26
killed on Jan. 9, 2023
in Guymon, Okla.

[See all victims](#)

Updated Jan. 11, 2023

[Read our methodology](#) → [View data on GitHub](#) [Submit a tip](#)

Chapter 2 of *Data Feminism* by Catherine D'Ignazio and Lauren Klein
<https://data-feminism.mitpress.mit.edu/>



2. Collect, Analyze, Imagine, Teach

Principle #2 of Data Feminism is to Challenge Power. Data feminism commits to challenging unequal power structures and working toward justice.

Chapter 1 of *Sorting Things Out* by
Geoffrey C. Bowker and Susan Leigh Star

Some Tricks of the Trade in Analyzing Classification



SORTING THINGS OUT
CLASSIFICATION AND ITS CONSEQUENCES
GEOFFREY C. BOWKER AND SUSAN LEIGH STAR

<https://direct.mit.edu/books/book/4738/Sorting-Things-OutClassification-and-Its>