



# Data Literacy

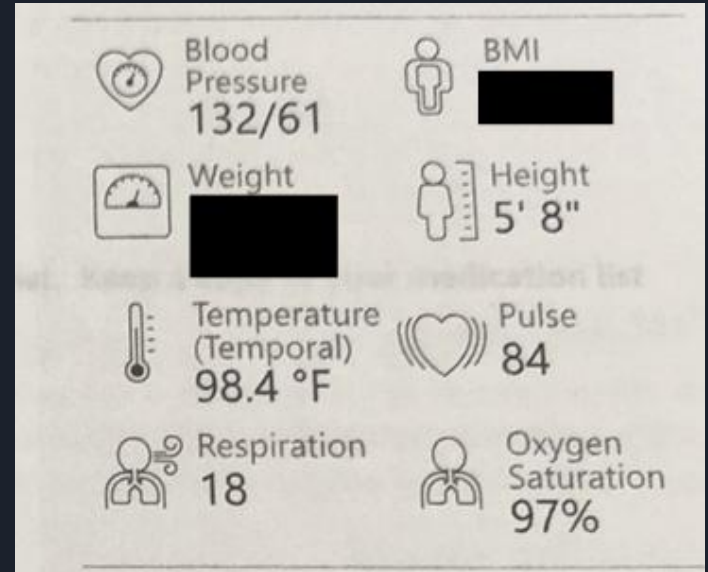
*...what is data, where is it,  
and why should we care?*

# Data Lit Agenda

- Discussion: Data in our Personal Lives
- Definitions
- The Stakes
- Activity: Uncovering Data Sources
- Data Myths
- Using Data in your Research
- Discussion: Thinking About & Using Data

# Data in Our Personal Lives

- What *is* data?
- What are some ways you created data today?
- What health data is collected on patients?
- Who owns all that data?



"After Visit Summary" - personal photo - Powered by [Epic](#)

# Data: Definitions

- DATA
  - “individual units of information”
    - (Shannon, 1948)
  - “building blocks of information”
    - (Kitchin, 2014)



*“Data Types” by João Batista Neto – [CC BY 3.0](#)*



# Data: Definitions

- BIG DATA
  - Many scholars still find it hard to define
  - But most agree on:
    - Volume + velocity
      - (Kitchin, 2014)
    - Complexity + technology
      - (Ward & Barker, 2013)



# Data: Definitions

- OPEN DATA

- “Open data is data that can be freely used, re-used, and redistributed by anyone as publicly available resources.”
  - (Adapted from [www.OpenDataHandbook.org](http://www.OpenDataHandbook.org))
- Open data aspires to be findable, accessible, interoperable, and reusable (FAIR).
  - (Adapted from [www.go-fair.org/fair-principles](http://www.go-fair.org/fair-principles))

# Data: The Stakes

- Privacy vs. public good
- Security vs. convenience
- Power
- Trust



*"Data Privacy" by PICOL – CC BY 3.0*

# Uncovering Data Sources

- [White House: the opioid epidemic cost \\$2.5 trillion over 4 years](#) | [Vox](#) | [German Lopez](#)

## White House: the opioid epidemic cost \$2.5 trillion over 4 years

Vox

The opioid epidemic is seriously hurting the US economy.

By German Lopez | @germanlopez | german.lopez@vox.com | Nov 1, 2019, 12:40pm EDT



Derrick Slaughter, 5, attends a march against the opioid epidemic in Norwalk, Ohio, on July 14, 2017. | Spencer Platt/Getty Images

The **opioid epidemic** cost the US \$696 billion in 2018 and more than \$2.5 trillion from 2015 to 2018, according to a new **estimate** by the White House Council of Economic Advisers.





# Uncovering Data Sources

- What data sources are referenced in this article?
- For you to better understand this article, what other data might have been helpful to know?
- Why do you think the overall costs of the opioid epidemic cited in this article conflict so much?

# Data Sources in the Article

- CDC Wonder Database:  
[wonder.cdc.gov](https://wonder.cdc.gov)

The screenshot displays the CDC WONDER website. At the top, there is a 'WONDER Search' bar with a search button. Below this is a 'WONDER Info' sidebar containing links for 'About CDC WONDER', 'What is WONDER?', 'Frequently Asked Questions', 'Data Use Restrictions', 'Data Collections', 'Citations', 'Republishing WONDER Data', and 'What's New?'. The main content area is titled 'CDC WONDER' and includes a description: 'WONDER online databases utilize a rich ad-hoc query system for the analysis of public health data. Reports and other query systems are also available.' Below this, there are three tabs: 'WONDER Systems', 'Topics', and 'A-Z Index'. The 'WONDER Systems' tab is active, showing a list of databases under the heading 'WONDER Online Databases'. These include 'AIDS Public Use Data', 'Births', 'Cancer Statistics', 'Environment' (with sub-links for Heat Wave Days, Daily Air Temperatures, Daily Land Surface Temperatures, Daily Fine Particulate Matter, Daily Sunlight, and Daily Precipitation), 'Mortality' (with sub-links for Underlying Cause of Death, Detailed Mortality, Compressed Mortality, US-Mexico Border Area Mortality, Multiple cause of death, Infant Deaths, Fetal Deaths, and Online Tuberculosis Information System), and 'Population' (with sub-links for Bridged-Race Population, Single-Race Population, Population Projections, Sexually Transmitted Disease Morbidity, and Vaccine Adverse Event Reporting). To the right of the 'WONDER Online Databases' list, there are two sections: 'Reports and References' (with links for Prevention Guidelines and Scientific Data and Documentation) and 'Other Query Systems' (with links for Healthy People 2010, NNDSS Annual Tables, NNDSS Weekly Tables, and 122 Cities Weekly Mortality).



# Data Sources in the Article

- CDC National Vital Statistics:  
[cdc.gov/nchs/nvss/index.htm](https://www.cdc.gov/nchs/nvss/index.htm)

## Key Mortality Statistics

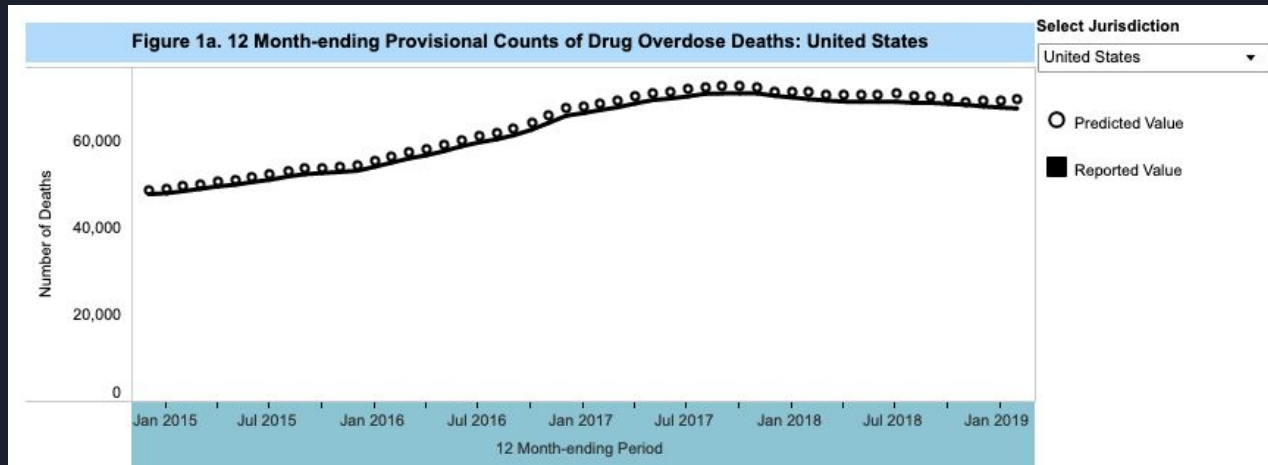
Data for United States in 2017

- Number of deaths: 2,813,503
- Death rate: 863.8 deaths per 100,000 population
- Life expectancy: 78.6 years
- Infant Mortality rate: 5.79 deaths per 1,000 live births

Source: [Mortality in the United States, 2017](#)   
[PDF – 505 KB]

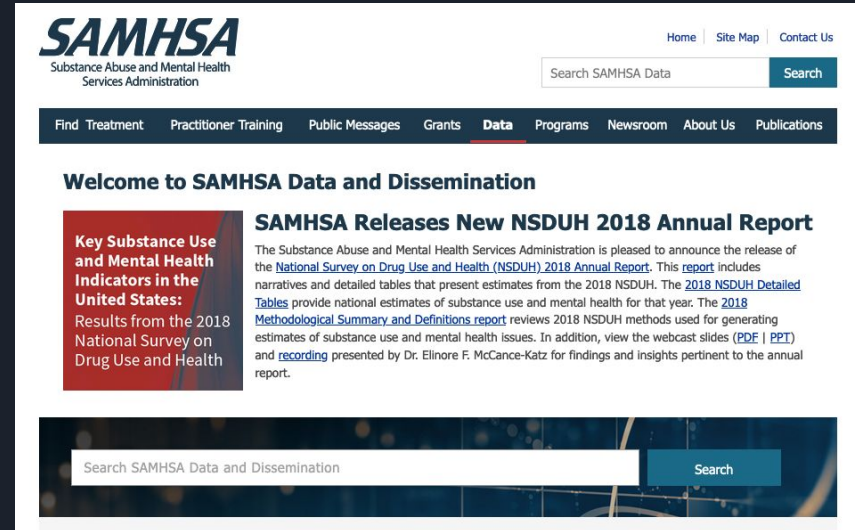
# Data Sources in the Article

- CDC National Vital Statistics | Provisional Drug Overdose Death Counts



# Data Sources in the Article

- DHHS Substance Abuse and Mental Health Services Administration: [samhsa.gov/data](https://www.samhsa.gov/data)



The screenshot shows the SAMHSA Data and Dissemination website. At the top, the SAMHSA logo is displayed with the text "Substance Abuse and Mental Health Services Administration". To the right of the logo are links for "Home", "Site Map", and "Contact Us". Below the logo is a search bar labeled "Search SAMHSA Data" with a "Search" button. A navigation bar below the search bar contains links for "Find Treatment", "Practitioner Training", "Public Messages", "Grants", "Data" (which is highlighted), "Programs", "Newsroom", "About Us", and "Publications". The main content area features a "Welcome to SAMHSA Data and Dissemination" heading. Below this, there is a red box with the text "Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health". To the right of this box is a section titled "SAMHSA Releases New NSDUH 2018 Annual Report" with a paragraph of text. At the bottom of the page, there is a large search bar labeled "Search SAMHSA Data and Dissemination" with a "Search" button.

**SAMHSA**  
Substance Abuse and Mental Health  
Services Administration

Home | Site Map | Contact Us

Search SAMHSA Data Search

Find Treatment Practitioner Training Public Messages Grants **Data** Programs Newsroom About Us Publications

**Welcome to SAMHSA Data and Dissemination**

**Key Substance Use and Mental Health Indicators in the United States:**  
Results from the 2018 National Survey on Drug Use and Health

**SAMHSA Releases New NSDUH 2018 Annual Report**

The Substance Abuse and Mental Health Services Administration is pleased to announce the release of the [National Survey on Drug Use and Health \(NSDUH\) 2018 Annual Report](#). This [report](#) includes narratives and detailed tables that present estimates from the 2018 NSDUH. The [2018 NSDUH Detailed Tables](#) provide national estimates of substance use and mental health for that year. The [2018 Methodological Summary and Definitions report](#) reviews 2018 NSDUH methods used for generating estimates of substance use and mental health issues. In addition, view the webcast slides ([PDF](#) | [PPT](#)) and [recording](#) presented by Dr. Elinore F. McCance-Katz for findings and insights pertinent to the annual report.

Search SAMHSA Data and Dissemination Search



# Data Myths

Data is not:

- Objective fact
- Free from bias
- (Just)  
computer-generated\*

It can be:

- Full of errors
- Manipulated
- Always  
human-influenced\*

*Essentially, data's just like any other source.*



# Using Data in your Research

- What data's often used for in research papers:
  - Supporting evidence for your argument
  - Benchmarking
  - Visualization
- What to avoid when using data:
  - Cherry-picking
  - Confirmation bias



# Using Data in your Research


- Where to find free, open data
  - Centers for Disease Control (CDC)
  - National Institutes of Health (NIH)
  - ICPSR
  - Data.gov
  - Data.WA.gov
  - ...and many more!



# Potential Data Sources

- NIH: Influenza Research Database: [fludb.org](https://fludb.org)

## Highlights



The map displays North America with various colored regions and numerous data points marked with pins. A large red area covers the eastern half of the continent, while green and yellow areas are in the west. Bird flyway lines are overlaid on the map. Navigation arrows are visible on the left and right sides of the map.

[Start Search](#)

[Tutorial](#)

### Surveillance

Human, avian and non-human mammalian surveillance data can be searched based upon location and various host characteristics for download and display on a map with bird flyway overlay. Create a report showing counts of surveillance records grouped by user-driven criteria.

#### Key Highlights:

- Human, avian and non-human mammalian surveillance data
- User-driven serotype comparison
- Display data on Google Map with flu prevalence color coded and bird flyways overlaid
- View results by metadata matrix report

# Potential Data Sources

- Institute for Social Research (ICPSR) - <https://www.icpsr.umich.edu/icpsrweb/ICPSR/>

The screenshot displays the ICPSR (Institute for Social Research) website interface. The header includes the ICPSR logo, the text 'Find & Analyze Data', and a 'Log In/Create Account' link. A navigation bar below the header contains links for 'FIND DATA', 'SEARCH/COMPARE VARIABLES', 'DATA-RELATED PUBLICATIONS', 'RESOURCES FOR STUDENTS', and 'HELP'.

The main content area is divided into two sections: 'Filters' on the left and 'Search Results' on the right.

**Filters:**

- Subject Terms:** A dropdown menu showing a list of terms: 'health care services (44)', 'medical care (32)', 'patients (28)', 'older adults (24)', and 'patient care (24)'. A 'view all' link is at the bottom.
- Restriction Type:** A dropdown menu.
- Data Format:** A dropdown menu.
- Collection Method:** A dropdown menu.

**Search Results:**

- Search Bar:** Contains the text 'respite care'. Buttons for 'GO' and 'VIEW ALL' are to the right. A 'search tips' link is below the bar.
- Results Summary:** Shows '124 results.' and a breakdown of results by type: 'Studies (124)', 'Variables (333)', 'Series (20)', 'Data-related Publications (1)', and 'ICPSR Website (0)'.
- Sort by:** A dropdown menu set to 'Study Relevance'.
- Navigation:** A set of buttons for navigating through the results, including '1', '2', '3', and arrows for previous/next.
- Table of Results:** A table with two columns: 'Study Title/Investigator' and 'Released/Updated'.

Study Title/Investigator	Released/Updated
1. <a href="#">Adult Respite Care Funded or Provided by State Governments, 1992: [United States]</a> (ICPSR 6584) Berdes, Celia <a href="#">more info</a>	2006-01-12
2. <a href="#">Chronic Illness and Caregiving, 2000: [United States]</a> (ICPSR 3402) Thamer, Mae <a href="#">more info</a>	2006-05-04
3. <a href="#">National Home and Hospice Care Survey, 1998</a> (ICPSR 3763) United States Department of Health and Human Services. National Center for Health Statistics	2006-01-18

# Potential Data Sources

- Partners in Information Access for the Public Health Workforce Data: [www.phpartners.org/public/health\\_stats](http://www.phpartners.org/public/health_stats)



The screenshot shows a web interface for 'Public Health Resources'. At the top, there's a dark blue header with the title 'Public Health Resources' in orange. Below this, a green square icon with a white bar chart and an upward arrow is next to the text 'Health Data, Tools & Statistics' in blue. Underneath, a link says 'Links to data on the health of a population.' in blue. There are two tabs: 'GENERAL' (highlighted with an orange border) and 'HEALTH TOPICS' (grey border). Below the tabs are two buttons: 'Open All' and 'Close All'. A list of data sources follows, each with an orange square icon containing a white plus sign: 'PUBLIC HEALTH WORKFORCE DATA', 'HEALTH DATA NEWS', 'COUNTY AND LOCAL HEALTH DATA', and 'STATE HEALTH DATA'.

**Public Health Resources**

**Health Data, Tools & Statistics**  
Links to data on the health of a population.

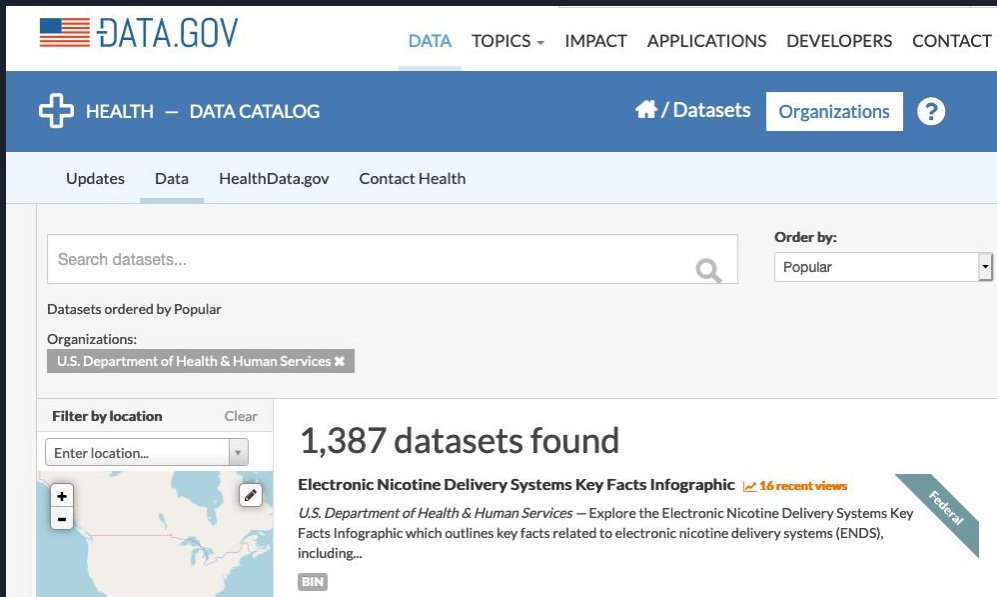
**GENERAL** **HEALTH TOPICS**

Open All Close All

- PUBLIC HEALTH WORKFORCE DATA**
- HEALTH DATA NEWS**
- COUNTY AND LOCAL HEALTH DATA**
- STATE HEALTH DATA**

# Potential Data Sources

- [www.data.gov](https://www.data.gov)



The screenshot displays the Data.gov website interface. At the top, the 'DATA.GOV' logo is visible alongside navigation links for DATA, TOPICS, IMPACT, APPLICATIONS, DEVELOPERS, and CONTACT. Below this, a blue header bar contains a plus icon, the text 'HEALTH - DATA CATALOG', and links for '/ Datasets', 'Organizations', and a help icon. A secondary navigation bar includes links for Updates, Data, HealthData.gov, and Contact Health. The main content area features a search bar with the placeholder 'Search datasets...', an 'Order by:' dropdown menu set to 'Popular', and a section for 'Organizations' listing the 'U.S. Department of Health & Human Services'. A 'Filter by location' section includes a 'Clear' button, a location input field, and a map of the United States. The search results section prominently displays '1,387 datasets found' and a featured result titled 'Electronic Nicotine Delivery Systems Key Facts Infographic' with 16 recent views. The description for this result mentions it is from the U.S. Department of Health & Human Services and explores key facts related to electronic nicotine delivery systems (ENDS). A 'Federal' label is visible on the right side of the featured result.



# Final Discussion

- Would your literature review benefit from a data source? If so, what might it be?
- By next Tuesday: Join the discussion on Canvas and answer at least one question there.



# References

Ahmad, F.B., Escobedo, L.A., Rossen, L.M., Spencer, M.R., Warner, M., & Sutton, P. (2019). Provisional drug overdose death counts. National Center for Health Statistics. Retrieved from <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>.

FAIR Principles. (n.d.) Retrieved from <https://www.go-fair.org/fair-principles/>.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage: Los Angeles.

Lopez, G. (2019). White House: the opioid epidemic cost \$2.5 trillion over 4 years. Vox. Retrieved from <https://www.vox.com/policy-and-politics/2019/11/1/20943599/opioid-epidemic-cost-white-house-economic-advisers>.



# References

Shannon, C. E. (1948). A mathematical theory of communication. Bell system technical journal, 27(3), 379-423.

Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

What Is Open Data?. (n.d.). Retrieved from <http://opendatahandbook.org/guide/en/what-is-open-data>.

**Note:** with the exception of the initial personal photo and website screenshots, images in this presentation are from the [Wikimedia Commons](#), an open access media resource.



# Databases

CDC Wonder: <https://wonder.cdc.gov/>

CDC National Vital Statistics System: <https://www.cdc.gov/nchs/nvss/index.htm>

Data.gov: <https://www.data.gov/>

Data.WA.gov: <https://data.wa.gov/>

DHHS Substance Abuse and Mental Health Services Administration:

<https://www.samhsa.gov/data/>

NIH Influenze Research Database: <https://www.fludb.org>

Inter-university Consortium for Political and Social Research (ICPSR):

<https://www.icpsr.umich.edu/icpsrweb/ICPSR/>

Partners in Information Access for the Public Health Workforce | Health Data, Tools,  
and Statistics: [https://www.phpartners.org/ph\\_public/health\\_stats](https://www.phpartners.org/ph_public/health_stats)





# Open Data Resources

- Open Data Institute | What is Open Data and Why Should We Care (with GIFs):  
<https://theodi.org/article/what-is-open-data-and-why-should-we-care/>
- World Bank | Open Data in 60 Seconds:  
<http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html>