

Maggie Mulvihill Associate Professor of
the Practice in Computational Journalism

Boston University

mmulvih@bu.edu

@maggieMulvihill

Extracting Data from PDF's: Another Way of “Scraping” Information



nefai 2021

Data Extraction v. Traditional Web Scrapping

- We'll be pulling valuable data out of a locked document
- We are “liberating” needed data from PDF's
- Data extraction from PDF's won't automatically update your data when the government agency updates their site
- Traditional web scrapping does give you that advantage



Autodetect Tables									
Clear All Selections									
Preview & Export Extracted Data									
Year	Month	Revenue	Expenses	Profit	Revenue	Expenses	Profit	Revenue	Expenses
2012	Jan	3,320,302	1,774	2,013	239	13,003	9,287	3,806	3,872
2012	Feb	3,320,324	2,528	2,643	154	11,418	8,622	2,966	2,966
2012	Mar	3,332,470	2,653	2,681	188	13,230	9,153	4,077	3,878
2012	Apr	3,332,348	2,655	2,485	70	11,927	8,610	3,317	3,378
2012	May	3,339,724	2,634	2,036	178	11,374	8,679	2,695	2,684
2012	Jun	3,342,028	2,768	2,703	65	11,968	9,134	2,834	2,805
2012	Jul	3,348,108	3,115	2,671	444	14,577	11,907	2,670	3,007
2012	Aug	3,348,205	3,253	2,476	753	16,346	12,176	3,869	4,664
2012	Sep	3,352,889	2,632	1,948	988	16,822	10,774	6,148	7,139
2012	Oct	3,360,009	3,013	3,148	165	16,752	11,662	7,613	7,115
2012	Nov	3,367,727	2,872	2,848	24	14,487	9,227	5,260	5,288
2012	Dec	3,373,015	4,245	3,724	621	9,254	7,932	1,662	2,287
2013	Jan	3,375,222	1,788	2,033	240	13,944	10,257	3,687	3,447
2013	Feb	3,375,669	2,441	2,703	272	11,781	8,566	3,265	2,933
2013	Mar	3,381,972	2,384	2,782	368	12,219	8,608	3,611	3,221
2013	Apr	3,388,793	2,619	3,103	124	12,987	9,980	3,969	3,851
2013	May	3,388,644	2,611	2,718	86	11,788	8,526	3,243	3,334
2013	Jun	3,395,978	2,719	2,657	148	11,335	9,376	2,016	2,362
2013	Jul	3,394,130	3,254	2,691	653	16,110	13,381	2,729	3,339
2013	Aug	3,395,489	3,111	2,660	467	16,351	12,972	3,266	3,616
2013	Sep	3,401,147	2,631	2,280	551	17,086	12,143	5,715	6,407
2013	Oct	3,405,814	3,387	2,676	722	16,976	12,627	6,784	7,477
2013	Nov	3,410,091	2,782	2,688	254	14,152	8,938	4,317	4,532
2013	Dec	3,419,623	4,387	3,838	952	10,388	9,786	1,642	2,288
2014	Jan	3,425,820	1,912	1,634	278	14,201	10,951	3,240	3,539
2014	Feb	3,425,368	2,716	2,536	171	11,974	9,938	2,186	2,582
2014	Mar	3,427,910	2,733	2,690	43	13,999	10,527	3,472	3,914
2014	Apr	3,432,622	2,896	2,460	476	12,864	10,160	2,526	3,237
2014	May	3,430,080	2,856	2,505	386	11,853	9,582	2,271	2,847
2014	Jun	3,432,919	3,280	2,162	1,078	11,807	9,916	1,891	2,449
2014	Jul	3,440,991	3,577	2,535	642	14,622	13,626	802	1,668
2014	Aug	3,442,448	2,681	2,675	16	16,176	12,496	2,609	3,144
2014	Sep	3,445,793	3,533	2,688	1,081	16,835	13,720	8,895	7,118
2014	Oct	3,448,111	3,322	2,879	443	16,263	12,966	8,674	8,488
2014	Nov	3,450,398	2,967	2,628	330	15,758	12,726	5,979	6,768
2014	Dec	3,456,164	4,495	4,274	221	12,375	9,347	3,028	3,469
2015	Jan	3,459,949	1,927	1,884	43	15,980	11,442	3,648	4,085
2015	Feb	3,475,034	2,646	2,651	46	13,487	11,214	2,273	2,665
2015	Mar	3,475,519	2,673	2,386	421	14,912	12,445	2,765	2,539
2015	Apr	3,478,558	2,711	2,913	352	13,279	10,600	2,470	2,608
2015	May	3,481,187	2,867	2,736	231	9,508	8,463	862	1,138
2015	Jun	3,482,492	3,713	3,492	282	12,721	12,121	2,700	2,452



Executive Department New Hires

**Data provided by the Human Resources Division*

Sometimes Government Makes It Easy to Get Data

Just Hit the Export or Download Button on a Government Site

Commission for the Blind	\$254,426.54
Commission for the Deaf and Hard Of Hearing	\$228,090.94
Department of Agricultural Resources	\$181,676.56
Department of Children & Families	\$14,466,090.68
Department of Conservation & Recreation	\$10,641,545.17
Department of Correction	\$3,651,954.38
Department of Developmental Services	\$13,350,376.02
Department of Early Education and Care	\$995,194.44
Department of Elder Affairs	\$177,620.20
Department of Elementary and Secondary Education	\$581,488.38

Total

ler Priv



Executive Depa

or a Department, Jok

DOWNLOAD



Data

We've made the data that powers Open Payroll available to you.

Get Data

Dataset	Description	View	Download
Payroll Transactions	Data for Payroll Transactions as displayed on this website.	View	Download

- But government sometimes “traps” data in PDF’s so we can’t access it easily to analyze
- We have to “free” the data so we can work with it. We use different tools to do that
- One way to do that is with **PDF converters** like a tool such as

cometdocs



Why Government Traps Data



“The crime stats are subject to being corruptible in an excel sheet. They have been changed in the past by persons unknown and this affects the veracity of the original data posted. If stats are posted on-line in a PDF format, this reduces the risk of contamination. [...] Effective immediately the stats should just be posted in a PDF format.”

-Minneapolis Police Department

<http://www.minnpost.com/data/2013/09/update-minneapolis-police-department-restores-accessible-data-format>

Vermont State Police - Officer Involved Shootings 1977 - present

An Official **Vermont** Government Website



Department of Public Safety
Vermont State Police



[COVID-19](#)
[Employment](#)
[Find A Station](#)
[Traffic Safety](#)
[Crime Prevention](#)
[Public Information](#)



JOIN OUR TEAM!

DPS Home

VSP Home

VSP COVID-19 response

Officer-involved shootings, 1977-present



Members of the Vermont State Police have been involved in 45 officer-involved shootings since 1977, when VSP recordkeeping on the subject began. In all 45 of the cases, the use of force was found to be justified following outside, independent reviews of the incidents. Information regarding these incidents is collected by the state police on a case-by-case basis and will be updated as needed. This page was most recently updated on April 13, 2020.

The data are provided in the following spreadsheet, which you may access and download in PDF format by clicking the image below:

[illegible]

Journalists Have to “Scrape” or Extract” Data from a PDF for Analysis

 **Massachusetts Bay Transportation Authority**

Getting Around ▾ Fares ▾ Contact Us ▾ More ▾  EN ▾ 

Home > MBTA Budgets and Financials > MBTA Budget

Financials

MBTA Budgets

Capital Investment Plan

Audited Financials

Major Bonds Programs

Fiscal and Management Control Board

MBTA Back on Track

Investor Information Disclaimer


MBTA Budget


Every year, the MBTA determines which activities and actions can be taken during the next fiscal year to continue to fund programs that provide safe, reliable, and accessible transportation services. These annual budgets are produced based on statutory guidelines, and can be downloaded below.


For MBTA financial statements produced in accordance with Generally Accepted Accounting Principles (GAAP), please visit our audited financial statements.


Note: The information in this section is presented subject to readers' acknowledgment of this disclaimer.

MBTA Payroll

 2017

 2015

 2014

 2013

Employee ID	Last Name	First Name	Title	Hire Date	Base Pay	Total 2017 Earnings
452912	Aalto	Joanna	AGM Capital Programs	03-Apr-16	\$185,000.00	\$28,461.55
72929	Aaron	Kavar	Laborer, Track	21-Sep-15	\$57,803.20	\$72,800.63
72021	Aaron	Keyena	Operator, P-T Surface	05-May-14	\$47,548.80	\$48,448.81
68101	Abany	Samuel	Sergeant	10-Apr-06	\$95,518.40	\$202,214.19
68415	Abbit	Cynthia	Inspector	05-Sep-06	\$83,304.00	\$98,700.28
9749	Abbott	Kenneth	Operator, Surface	03-Feb-92	\$74,588.80	\$85,631.68
72103	Abbott	Charles	Operator, Surface	23-Jun-14	\$63,398.40	\$74,230.46
73625	Abbott	Corey	Operator, P-T Surface	09-Jan-17	\$30,763.20	\$29,608.46
451427	Abdelkarim	Loay	Mgr Design	27-Feb-12	\$117,279.99	\$116,354.38
71810	Abdella	Nasir	Operator, Surface	06-Jan-14	\$67,121.60	\$73,542.99
72734	Abdelouahhab	Ali	Operator, Surface	22-Jun-15	\$63,398.40	\$63,624.21
67965	Abdelrahman	Ashraf	Supv Trans - Bus	11-Oct-05	\$91,312.90	\$89,721.43
72691	Abderazzak	Sammy	Auditor, Field	01-Jun-15	\$65,880.52	\$64,723.80
			Supv Transm&Dist			
466160	Abdulakhatov	Akmal	Eng&Sfty	12-Feb-17	\$111,152.29	\$94,330.77
68620	Abdullahi	Abdul-Hakim	Trackperson	10-Oct-06	\$71,552.00	\$93,345.76
69697	Abell	Paul	Worker, Sheet Metal	12-Aug-13	\$96,553.60	\$108,503.17
67679	Abernathy	Sandra	Motorperson, Rapid Transit	04-Apr-05	\$77,563.20	\$55,099.48
69204	Abibi	Abdelmalek	Operator, Surface	10-Sep-07	\$74,588.80	\$92,339.26
66635	Abkarian	Ropen	Operator, Surface	15-Oct-02	\$74,588.80	\$15,562.41
67420	Abraham	Bernard	Operator, Surface	07-Sep-04	\$74,588.80	\$110,055.70
369115	Abrahamson	Daniel	Spec, Energy Conservation	06-Jun-16	\$71,370.43	\$70,117.26
73316	Abramo	Michael	Chief Financial Officer	06-Jun-16	\$210,000.00	\$186,635.50
67817	Abreu	Estarlo	Operator, Surface	11-Jul-05	\$74,588.80	\$87,679.32
71537	Abreu	Yris	Operator, Surface	20-May-13	\$74,588.80	\$58,391.87
73503	Abujade	Olalekan	Operator, P-T Surface	06-Sep-16	\$41,964.00	\$41,757.36
477676	Acchat	Kerby	Operator, P-T Surface	19-Nov-17	\$30,763.20	\$2,200.00
71941	Accime	Susie	Operator, Surface	24-Feb-14	\$67,121.60	\$97,051.46
67909	Acevedo	Tyrone	Repairer, Rail	03-Oct-05	\$86,715.20	\$129,078.87
72828	Acharya	Horn	Operator, Surface	20-Jul-15	\$55,952.00	\$44,999.50
92047	Adams	Lee	Worker, Sheet Metal	22-Aug-00	\$96,553.60	\$144,456.73
7796	Adams	Edward	Inspector, Chief - SL	13-Feb-84	\$87,214.40	\$130,845.02
67527	Adams	Jeffrey	Repairer, Rail	13-Dec-04	\$86,715.20	\$119,303.28
66003	Adams	Sheila	Dispatcher Bus	09-Jul-17	\$92,934.40	\$117,796.60

JUVENILE					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	2,842	66,879	3,965	6,085	107,665
MALE	2,245	58,838	2,837	5,141	89,702
UNK	0	10	2	5	14
TOTAL	5,087	125,727	6,804	11,231	197,381

INVOLUNTARY					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	107	2,801	102	523	2,789
MALE	108	2,948	78	521	3,258
UNK	0	0	0	0	1
TOTAL	215	5,749	180	1,044	6,048

CATASTROPHE VICTIM					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	2	10	0	6	93
MALE	6	25	1	7	129
UNK	0	0	0	0	0
TOTAL	8	35	1	13	222

LEGEND	
JUV	Juvenile
END	Endangered
INV	Involuntary
DIS	Disability
CAT	Catastrophe Victim

ENDANGERED					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	374	4,194	356	685	11,703
MALE	494	5,596	347	950	16,968
UNK	0	6	0	14	6
TOTAL	868	9,796	703	1,649	28,677

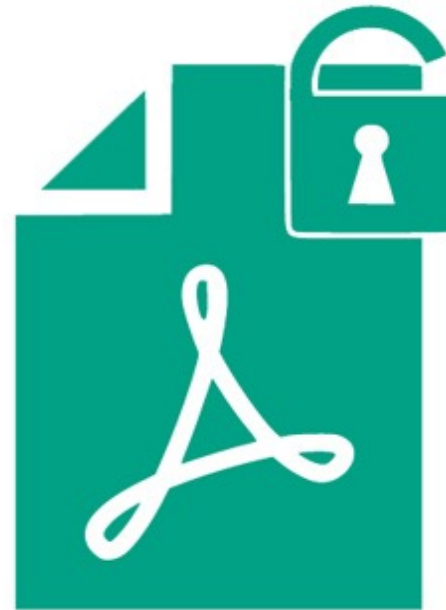
DISABILITY					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	335	4,383	107	194	6,445
MALE	425	6,408	150	394	10,356
UNK	0	1	0	4	2
TOTAL	760	10,792	257	592	16,803

OTHER					
	ASIAN	BLACK	INDIAN	UNK	WHITE*
FEMALE	1,629	12,066	765	1,445	30,334
MALE	2,209	18,381	863	2,284	42,388
UNK	0	2	2	2	6

2020 National Crime Information Center (NCIC) Missing Person and Unidentified Person Statistics Pursuant to the Requirements of the Crime Control Act of 1990, Pub. L. No. 101-647, 104 Stat. 4789

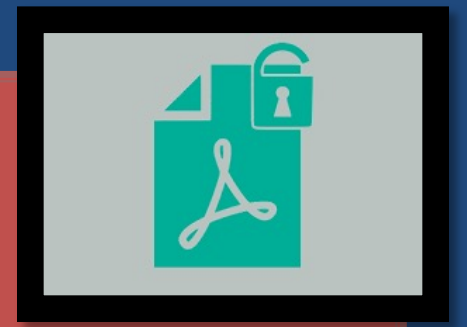
Interesting tables to extract; This is missing person data that you have as a PDF in the folders I sent you

A free tool like
Tabula comes in
handy



TABULA
liberate data tables trapped inside PDFs

What Is TABULA?



- Free extraction tool to turn tabular data locked in PDFs into exportable CSV or spreadsheet files
- Works on Windows, MAC or Linux systems
- Works on text-based PDF's, where you can select text
- Not scanned documents

Differences: Text Based and Scanned

- Text-based files
- PDF stores exact positions of every character
- TABULA can take those positions and extract the characters

- Scanned PDF's are a collection of images
- Can't extract the characters
- Could use OCR but accuracy can be problem



Install TABULA

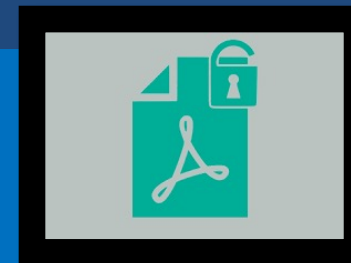


- **Download & Install Tabula**

- Windows & Linux users will need a copy of [Java](#) installed. You can [download Java here](#). (Java is included in the Mac version.)
- Download the version of Tabula for your operating system:
 - **Windows:** [tabula-win.zip](#)
 - **Mac OS X:** [tabula-mac.zip](#)
 - **Linux/Other:** [tabula-jar.zip](#), view README.txt inside for instructions

- Extract the zip file. (Instructions: [Windows](#), [Mac](#))
- Go into the folder you just extracted. Run the "Tabula" program inside.
- A web browser will open. If it doesn't, open your web browser, and go to <http://localhost:8080>. There's Tabula!

Launching TABULA



```
mysql> use elephant1
Database changed
mysql> SOURCE C:\xampp\htdocs\elephant12\db\mas_gis_city_ip.sql;
ERROR:
Unknown command '\x'.

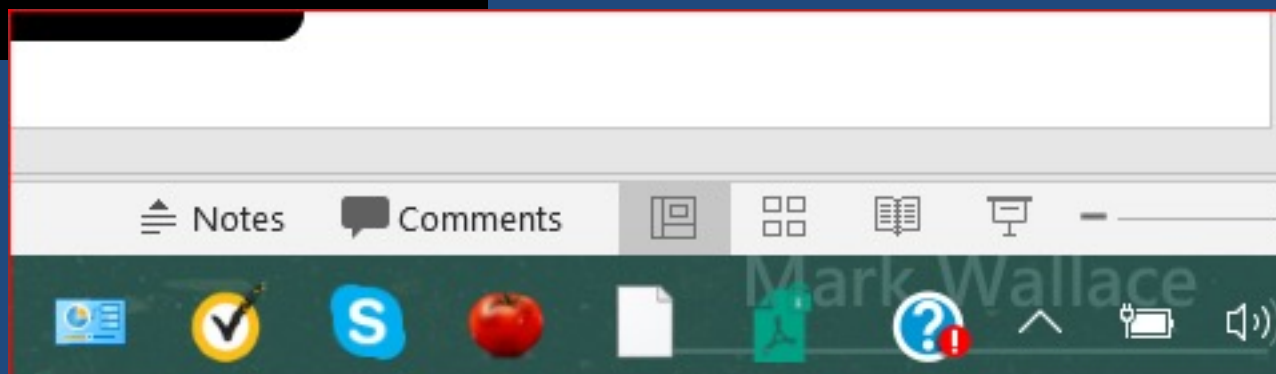
For information about MySQL products and services, visit:
  http://www.mysql.com/
For developer information, including the MySQL Reference Manual, visit:
  http://dev.mysql.com/
To buy MySQL Enterprise support, training, or other products, visit:
  https://shop.mysql.com/

List of all MySQL commands:
Note that all text commands must be first on line and end with ';'
?      (?)  Synonym for 'help'.
clear  (C)  Clear the current input statement.
connect (r)  Reconnect to the server. Optional arguments are db and host.
delimiter (d) Set statement delimiter.
ego    (G)  Send command to mysql server, display result vertically.
exit   (q)  Exit mysql. Same as quit.
go     (g)  Send command to mysql server.
help   (h)  Display this help.
notee  (t)  Don't write into outfile.
print  (p)  Print current command.
prompt (R)  Change your mysql prompt.
quit   (q)  Quit mysql.
rehash (#) Rebuild completion hash.
source (s)  Execute an SQL script file. Takes a file name as an argument.
status (s)  Get status information from the server.
tee     (T)  Set outfile [to_outfile]. Append everything into given outfile.
use     (U)  Use another database. Takes database name as argument.
charset (C) Switch to another charset. Might be needed for processing binlog with multi-byte charsets.
warnings (W) Show warnings after every statement.
nowarning (w) Don't show warnings after every statement.

For server side help, type 'help contents'

->
```

- ❑ Once you launch TABULA it will first run on the command
- ❑ Pin it to your favorites or taskbar for easy launching



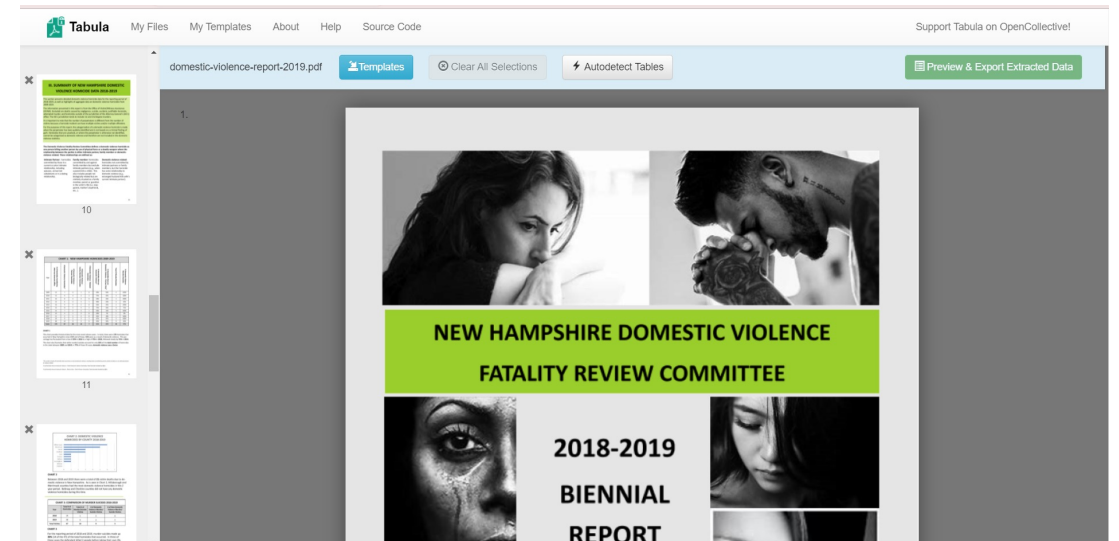
Importing Data



- I sent Justin a folder of government data in all six New England states in filetype:pdf format
- Let's liberate data from one table together and then you pick one from your state that you can run through Tabula to extract the information you want
- Go to your "NEFAI 2021 PDFS"

New Hampshire Domestic Violence Report

- You'll See Your State Folders
- In the NH folder, there is a PDF report
- **You'll also see a .PDF file called New Hampshire Domestic Violence Fatality Review Committee.**
- Download the report to your desktop and save it clearly marked

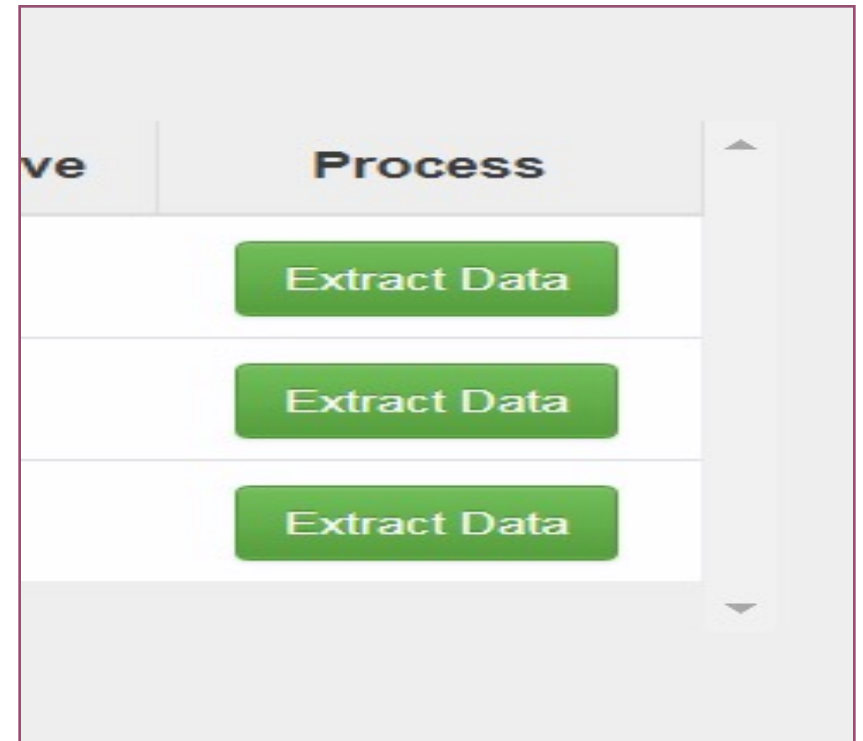


Liberating Data With Tabula

Extract Your Data

Extract Your Data

- Click on the green “extract data” button to open your PDF



1.



**NEW HAMPSHIRE DOMESTIC VIOLENCE
FATALITY REVIEW COMMITTEE**



III. SUMMARY OF NEW HAMPSHIRE DOMESTIC VIOLENCE HOMICIDE DATA 2018-2019

its actual members divided domestic officers' raw data by the "sampling area" of 2008–2010 and it also highlights aggregate data on domestic violence homicides from 2008–2010.

The information presented in this report is from the Office of Statistics/Information Systems, primarily because data created by agencies, such as, coroners, available for release, are generally not as authoritative as the authoritative nature of the information of the Attorney General's (AG) office. The AG's jurisdiction must include law enforcement and legal inquiries.

It is important to note that the number of populations is different from the number of officers because a homicide murder case can include many units and/or multiple officers.

For purposes of this report, the subgroup data of a currently assigned homicide is made available for release to law enforcement and is contained on a continuing basis of all. Homicides that are unsolved, or where the potential for an investigation is declined, cannot be assigned to domestic violence and therefore are not included in this domestic violence analysis.

Family members include family members defined as domestic violence homicide victims' first-degree killing partner or in-laws (as far as a family member when the relationship was not the victim's) or spouse, family member or domestic violence victim. These categories are defined as:

Immediate family members Domestic violence homicide victims who have a spouse or other immediate family member, including spouse, domestic violence victim or a sibling immediately.	Family member Domestic violence homicide victims who are controlled by and against family members but exclude spouse or other immediate family member (e.g., when a parent(s) is a DVH). This also includes spouse or other immediate family member who is not a victim, spouse or a family member, present or possible in the victim's life (e.g., step-parent, mother-in-law/step-mother, etc.).	Domestic violence victim Domestic violence homicide victims' partner or family members. In other words, this has been identified as a family member in a DVH.
--	--	---

10

Chart 1. NEW HAMPSHIRE TOWNERS 2009-2019

Variable	Mean	SD	Min	Max
Age (years)	38.5	12.5	18	65
Gender (Male/Female)	50/50	-	-	-
Marital Status (Married/Single)	60/40	-	-	-
Education Level (High School/College/Graduate)	30/40/30	-	-	-
Occupation (Student/Professional/Unemployed)	20/40/40	-	-	-
Family Income (Low/Medium/High)	30/40/30	-	-	-
Health Status (Good/Fair/Poor)	50/30/20	-	-	-
Smoking Status (Smoker/Non-smoker)	20/80	-	-	-
Alcohol Consumption (Regular/Occasional/None)	10/40/50	-	-	-
Exercise Frequency (Daily/Weekly/Monthly/None)	20/30/30/20	-	-	-
Stress Level (Low/Medium/High)	30/40/30	-	-	-
Family Size (1-2/3-4/5+)	20/40/40	-	-	-
Health Insurance (Yes/No)	80/20	-	-	-
Chronic Conditions (Hypertension/Diabetes/None)	10/10/80	-	-	-
Medication Use (Regular/Irrregular/None)	20/30/50	-	-	-
Healthcare Seeking Behavior (Regularly/When Needed/Not at All)	30/40/30	-	-	-
Healthcare Access (Easy/Difficult/None)	40/30/30	-	-	-
Healthcare Costs (Low/Medium/High)	30/40/30	-	-	-
Healthcare Satisfaction (Satisfied/Dissatisfied/Neutral)	40/30/30	-	-	-
Healthcare Quality (Good/Fair/Poor)	50/30/20	-	-	-
Healthcare Accessibility (Easy/Difficult/None)	40/30/30	-	-	-
Healthcare Affordability (Affordable/Unaffordable/Neutral)	30/40/30	-	-	-
Healthcare Effectiveness (Effective/Ineffective/Neutral)	40/30/30	-	-	-
Healthcare Safety (Safe/Unsafe/Neutral)	50/30/20	-	-	-
Healthcare Transparency (Transparent/Not Transparent/Neutral)	40/30/30	-	-	-
Healthcare Accountability (Accountable/Not Accountable/Neutral)	40/30/30	-	-	-
Healthcare Communication (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Collaboration (Collaborative/Not Collaborative/Neutral)	40/30/30	-	-	-
Healthcare Innovation (Innovative/Not Innovative/Neutral)	30/40/30	-	-	-
Healthcare Research (Active/Inactive/Neutral)	40/30/30	-	-	-
Healthcare Education (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Training (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Certification (Certified/Not Certified/Neutral)	40/30/30	-	-	-
Healthcare Accreditation (Accredited/Not Accredited/Neutral)	40/30/30	-	-	-
Healthcare Regulation (Regulated/Not Regulated/Neutral)	40/30/30	-	-	-
Healthcare Monitoring (Monitored/Not Monitored/Neutral)	40/30/30	-	-	-
Healthcare Evaluation (Evaluated/Not Evaluated/Neutral)	40/30/30	-	-	-
Healthcare Improvement (Improved/Not Improved/Neutral)	40/30/30	-	-	-
Healthcare Innovation (Innovative/Not Innovative/Neutral)	30/40/30	-	-	-
Healthcare Research (Active/Inactive/Neutral)	40/30/30	-	-	-
Healthcare Education (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Training (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Certification (Certified/Not Certified/Neutral)	40/30/30	-	-	-
Healthcare Accreditation (Accredited/Not Accredited/Neutral)	40/30/30	-	-	-
Healthcare Regulation (Regulated/Not Regulated/Neutral)	40/30/30	-	-	-
Healthcare Monitoring (Monitored/Not Monitored/Neutral)	40/30/30	-	-	-
Healthcare Evaluation (Evaluated/Not Evaluated/Neutral)	40/30/30	-	-	-
Healthcare Improvement (Improved/Not Improved/Neutral)	40/30/30	-	-	-
Healthcare Innovation (Innovative/Not Innovative/Neutral)	30/40/30	-	-	-
Healthcare Research (Active/Inactive/Neutral)	40/30/30	-	-	-
Healthcare Education (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Training (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Certification (Certified/Not Certified/Neutral)	40/30/30	-	-	-
Healthcare Accreditation (Accredited/Not Accredited/Neutral)	40/30/30	-	-	-
Healthcare Regulation (Regulated/Not Regulated/Neutral)	40/30/30	-	-	-
Healthcare Monitoring (Monitored/Not Monitored/Neutral)	40/30/30	-	-	-
Healthcare Evaluation (Evaluated/Not Evaluated/Neutral)	40/30/30	-	-	-
Healthcare Improvement (Improved/Not Improved/Neutral)	40/30/30	-	-	-
Healthcare Innovation (Innovative/Not Innovative/Neutral)	30/40/30	-	-	-
Healthcare Research (Active/Inactive/Neutral)	40/30/30	-	-	-
Healthcare Education (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Training (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Certification (Certified/Not Certified/Neutral)	40/30/30	-	-	-
Healthcare Accreditation (Accredited/Not Accredited/Neutral)	40/30/30	-	-	-
Healthcare Regulation (Regulated/Not Regulated/Neutral)	40/30/30	-	-	-
Healthcare Monitoring (Monitored/Not Monitored/Neutral)	40/30/30	-	-	-
Healthcare Evaluation (Evaluated/Not Evaluated/Neutral)	40/30/30	-	-	-
Healthcare Improvement (Improved/Not Improved/Neutral)	40/30/30	-	-	-
Healthcare Innovation (Innovative/Not Innovative/Neutral)	30/40/30	-	-	-
Healthcare Research (Active/Inactive/Neutral)	40/30/30	-	-	-
Healthcare Education (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Training (Good/Poor/Neutral)	50/30/20	-	-	-
Healthcare Certification (Certified/Not Certified/Neutral)	40/30/30	-	-	-

Abstract Available from the Internet for the most recent election year, in total there were 176 territories that submitted New Hampshire state 500 and of those, 938 were as a result of domestic delivery. The per capita income for the state rose from a low of \$20,000 in 2004 to a high of \$25,000 in 2008, followed closely by \$26 in 2009. The state also reported that public member numbers rose for only 30% of the total number of territories in the state between 2005 and 2009, or 11% of those 30 years. Domestic delivery was a factor.

The number sample size depends on the accuracy of the estimated effect sizes. Smaller effect sizes require larger samples.

11

CHART 2: DOMESTIC VIOLENCE

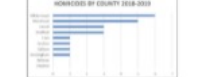


CHART 2

Between 2008 and 2010 there were a total of 21 victim deaths due to domestic violence in New Hampshire. As is seen in Chart 3, 14 (67 percent) and 10 (48 percent) counties had the most domestic violence homicides in this 2 year period. Belknap and Cheshire counties did not have any domestic violence homicides during this time.

Year	Total # of Murders	Rate of Murders (murders per 100,000)	# of Domestic Murders (murders per 100,000)	# of Non-domestic Murders (murders per 100,000)
2010	124	3.2	2.1	1.1
2011	120	3.1	2.0	1.1
2012	122	3.2	2.1	1.1

CHART 3

For the reporting period of 2018 and 2019, murder victims made up 88% (34 of 47) of the total homicides that occurred. In three of these cases the defendant killed 2 people before taking their own life.

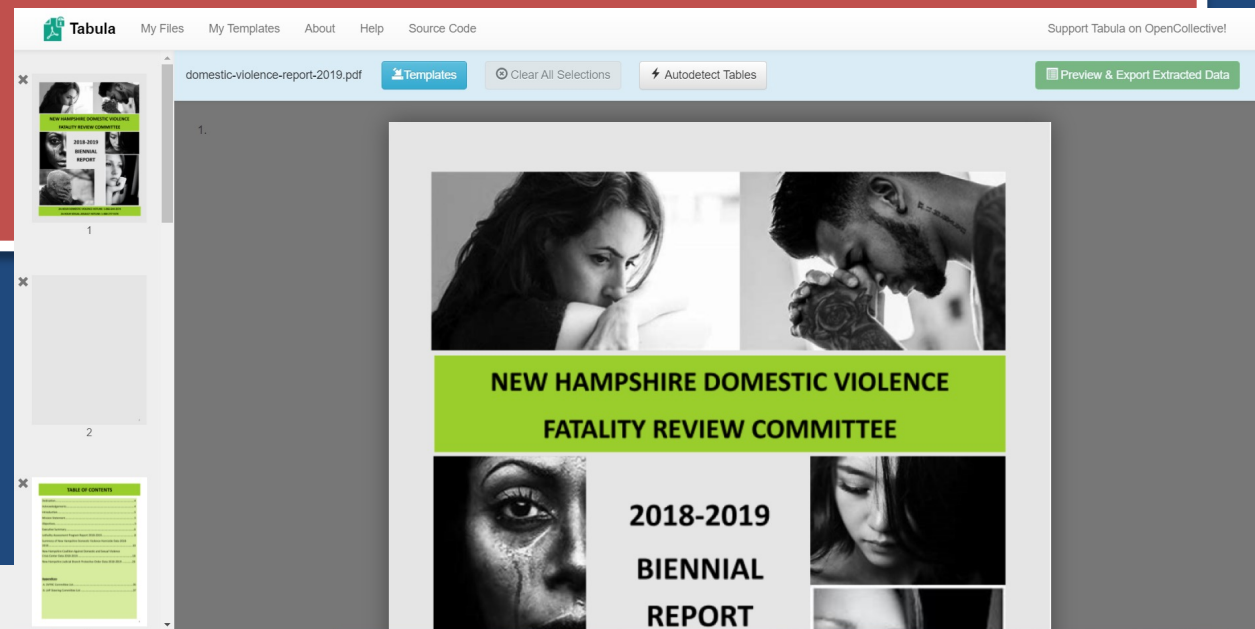
Browse and Find File

Import one or more PDFs

Browse...

domestic-violence-report-2019.pdf

Import to Load File



Import one or more PDFs

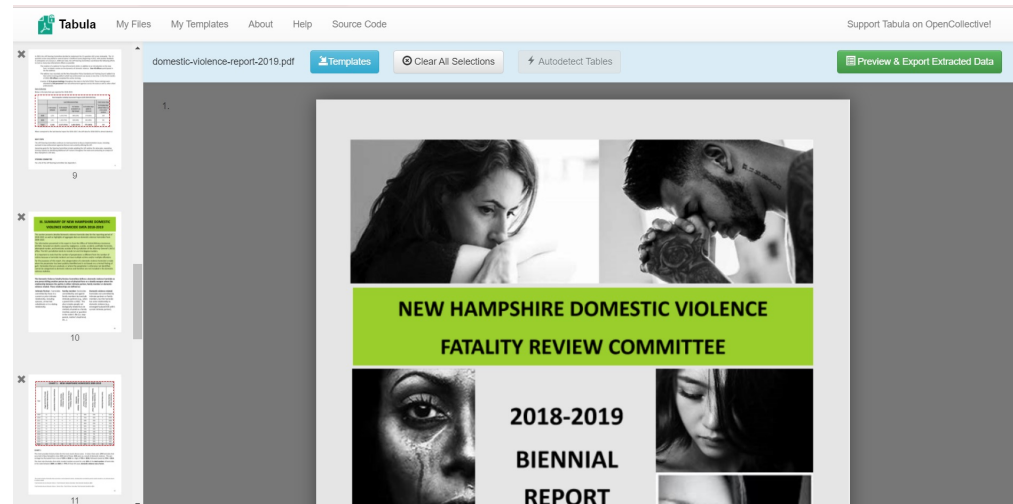
Browse...

domestic-violence-report-2019.pdf


Import

Auto Detect Tables

- If “auto detect” is running, wait until the little lightning bolt stops moving
- Then you can click it
- It will start to run through your data and highlight in pink anything it detects is in tabular format, - columns and rows
- Once it is finished, it will stop, and you can export your data or revise your selections



Once Imported, You Can “Extract Data”

 **Tabula** [My Files](#) [My Templates](#) [About](#) [Help](#) [Source Code](#) [Support Tabula on OpenCollective!](#)

Import one or more PDFs

[Browse...](#) [Import](#)

Imported PDFs

File Name	Size	Pages	Date Added	Remove	Process
domestic-violence-report-2019.pdf	2429 kB	38	04 Nov 2021 12:14	✕	Extract Data
Microsoft Word - Annual Report 2020.pdf	613 kB	10	04 Nov 2021 02:05	✕	Extract Data
Federal Firearms Licensee (FFL) Theft_Loss Report 2017.pdf	534 kB	6	04 Nov 2021 01:38	✕	Extract Data
Federal Firearms Licensee (FFL) Theft_Loss Report 2017.pdf	534 kB	6	02 Nov 2021 15:59	✕	Extract Data
overdose-deaths-by-county-including-map-may-2017.pdf	402 kB	3	02 Nov 2021 14:31	✕	Extract Data
4thQtr2013 final.pdf	1510 kB	62	24 Sep 2021	✕	Extract Data

Previewing



- TABULA will generate images you can PREVIEW
- Scroll through what it has imported to make sure it all correct
- You can see the tabular data within each page
- This is what TABULA will go in and extract



III. SUMMARY OF NEW HAMPSHIRE DOMESTIC VIOLENCE HOMICIDE DATA 2018-2019

This section presents detailed domestic violence homicide data for the reporting period of 2018-2019, as well as highlights of aggregate data on domestic violence homicides from 2008-2019.

The information presented in this report is from the Office of Justice Services, Domestic Violence Unit, which is the state's central repository for domestic violence data. The data is compiled from reports filed by law enforcement agencies throughout the state, as well as from the records of the Attorney General's Office.

It is important to note that the number of homicides is different from the number of victims because a homicide victim may be killed by multiple offenders, or a single offender may be responsible for multiple homicides.

For the purposes of this report, the definition of a domestic violence homicide is a homicide where the perpetrator has been publicly identified as a victim of a criminal act of domestic violence, or where the perpetrator is otherwise not identified, but the victim is identified as a domestic violence victim.

The Domestic Violence Fatality Review Committee defines a domestic violence homicide as the killing of a victim by a perpetrator who is a family member, intimate partner, or someone with whom the victim has a close relationship.

Intimate Partner Homicide: Homicide committed by a family member or intimate partner of the victim. This includes homicide committed by a spouse, partner, or former spouse/partner.

Family Member Homicide: Homicide committed by a family member of the victim, other than an intimate partner. This includes homicide committed by a parent, child, sibling, or other family member.

Domestic Violence Homicide: Homicide committed by a family member or intimate partner of the victim, or by a family member of the victim, other than an intimate partner.

10



CHART 1: NEW HAMPSHIRE DOMESTIC VIOLENCE HOMICIDE DATA 2018-2019

Category	2018	2019	Total
Intimate Partner Homicide	1	1	2
Family Member Homicide	1	1	2
Domestic Violence Homicide	2	2	4
Total	2	2	4

CHART 1: This chart provides information on the number of domestic violence homicides in New Hampshire for the years 2018 and 2019. The data shows that there were two domestic violence homicides in 2018 and two in 2019, for a total of four.

Domestic Violence Homicide: Homicide committed by a family member or intimate partner of the victim, or by a family member of the victim, other than an intimate partner.

11



CHART 2: DOMESTIC VIOLENCE HOMICIDES BY COUNTY 2018-2019

County	2018	2019	Total
Rockingham	1	1	2
Sherburne	1	1	2
Total	2	2	4

CHART 2: Between 2018 and 2019, there were a total of four domestic violence homicides in New Hampshire. Two of these were in Rockingham County and two were in Sherburne County.

CHART 3: COMPARISON OF DOMESTIC VIOLENCE HOMICIDES 2018-2019

Category	2018	2019	Total
Intimate Partner Homicide	1	1	2
Family Member Homicide	1	1	2
Domestic Violence Homicide	2	2	4
Total	2	2	4

1.



NEW HAMPSHIRE DOMESTIC VIOLENCE FATALITY REVIEW COMMITTEE



2018-2019 BIENNIAL REPORT

A series of **15 in person trainings** throughout the state in the fall of 2019. These trainings were attended by **245 personnel** from law enforcement agencies across the state as well as other allied professionals.

Data Collection

Below is the data that was reported for 2018-2019:

There are lots of interesting tables we might want to extract and have in a spreadsheet format

New Hampshire Lethality Assessment Program (LAP) 2018-2019 Data					
	Law Enforcement Data				Crisis Center Data
	# of screens initiated	# of screens completed	# of victims screened in as high danger	# of victims that spoke to advocate	# of victims that utilized follow-up crisis center services
2018	2,255	1,635 (73%)	864 (53%)	374 (43%)	269
2019	1,911	1,502 (79%)	818 (54%)	401 (49%)	291
Total	4,166	3,137 (75%)	1,682 (54%)	775 (46%)	560

Finding for a Story: Steady Increase in Domestic Violence Murder/Suicides

CHART 1: NEW HAMPSHIRE HOMICIDES 2009-2019

Year	# of Homicides Handled by Attorney Generals Office	# Domestic Violence Homicides ¹	# Domestic Violence - Partner Homicides	# Domestic Violence - Family Member Homicides	# Domestic Violence - Related Homicides	% Homicides that are Domestic Violence ²	% Homicides that are Domestic Violence - Partner Only ³	Total # Murder/Suicides	% Domestic Violence Murder/Suicides
2009	12	7	6	1	0	58%	50%	4	100%
2010	12	9	5	2	2	75%	42%	3	100%
2011	16	8	4	4	0	50%	25%	2	100%
2012	13	6	3	2	1	46%	23%	6	33%
2013	16	9	4	5	0	56%	25%	2	100%
2014	15	11	5	4	2	73%	33%	8	75%
2015	14	7	2	4	1	50%	14%	4	100%
2016	14	4	1	3	0	29%	7%	1	0%
2017	11	5	2	3	0	45%	18%	1	100%
2018	19	9	5	3	1	47%	47%	5	100%
2019	28	12	9	3	0	43%	32%	9	44%
Totals	170	87	46	34	7	51%	27%	45	77%

Click on “Preview & Export Extracted Data”



Tabula My Files My Templates About Help Source Code Support Tabula on OpenCollective!

domestic-violence-report-2019.pdf Templates Clear All Selections Autodetect Tables Preview & Export Extracted Data

1.

10

11

NEW HAMPSHIRE DOMESTIC VIOLENCE
FATALITY REVIEW COMMITTEE

2018-2019
BIENNIAL
REPORT

Your Data Will Export Into a Spreadsheet “View”

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

← Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

 Stream

 Lattice

Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

Still look wrong?

[Contact the developers](#) and

domestic-violence-report-2019.pdf

Export Format:

CSV

Export

Copy to Clipboard

CHART 3: COMPARISON OF MURDER SUICIDES 2018-2019

Year	Total # of Homicides	Total # of Murder/Suicide Victims	# of Domestic Violence Murder/ Suicide Victims	# of Non-Domestic Violence Murder/ Suicide Victims
2018	19	5	5	0
2019	28	9	4	5

CHART 5: HOMICIDE CAUSE OF DEATH 2018-2019

Cause of Death	Domestic Violence Homicide Victims	Non-Domestic Violence Homicide Victims
Firearm	10	15
Hand Gun	9	15
Long Gun	1	0
Cut/Stab	4	4
Blunt Impact	3	5
Strangulation	2	0
Other	2	2

You May Have To Do Some Cleaning

domestic-violence-report-2019.pdf Export Format: CSV Export Copy to Clipboard				
	3Domestic Violence - Partner Only			
	% Homicides that are			
	2Domestic Violence			
2009	12 7 6	1 0 58% 50%	4	100%
2010	# D1o2m estic Vio9le nce - Rela5t ed	2 2 75% 42%	3	100%
	Homicides			
2011	16 8 4	4 0 50% 25%	2	
2012	13 6 3	2 1 46% 23%	6	
	# Domestic Violence -			
2013	F1a6m ily Memb9e r Homicid4e s	5 0 56% 25%	2	100%
2014	15 11 5	4 2 73% 33%	8	75%
2015	14 7 2	4 1 50% 14%	4	100%
	# Domestic Violence -			
2016	14 Partner H4o micides 1	3 0 29% 7%	1	0%
2017	11 5 2	3 0 45% 18%	1	100%
2018	19 9 5	3 1 47% 47%	5	100%
	1			

Use Open Refine – free software on

Or You Can
Back To Your
“Extraction” and
Delete Data You
Don’t Want

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

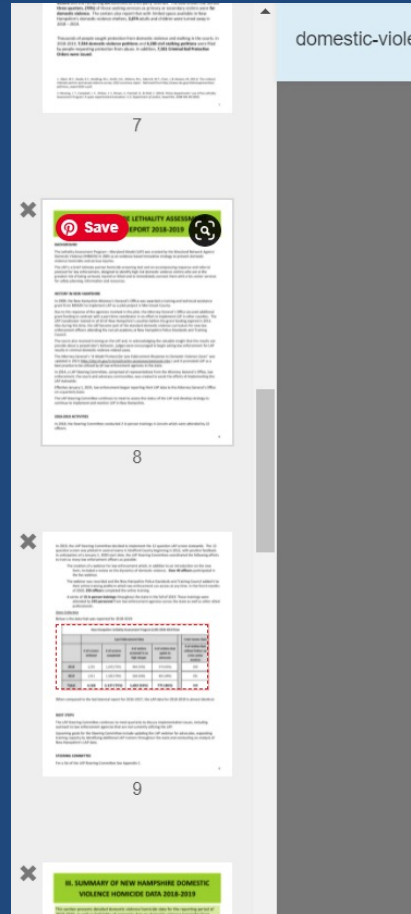
Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

← Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the

Revise Your Selection If You Don't Want All These Tables



- Just click the X to the upper left of each page
- Tabula will “deselect” the table
- Repeat “Preview & Export Selected Data”

[My Files](#)[My Templates](#)[About](#)[Help](#)[Source Code](#)[Support Tabula on OpenCollective](#)

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

[← Revise selection\(s\)](#)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice**

domestic-violence-report-2019.pdf

Export Format:

CSV

[⬇ Export](#)[📎 Copy to Clipboard](#)

CHART 3: COMPARISON OF MURDER SUICIDES 2018-2019

Year	Total # of Homicides	Total # of Murder/Suicide Victims	# of Domestic Violence Murder/Suicide Victims	# of Non-Domestic Violence Murder/Suicide Victims
2018	19	5	5	0
2019	28	9	4	5

CHART 5: HOMICIDE CAUSE OF DEATH 2018-2019

Cause of Death	Domestic Violence Homicide Victims	Non-Domestic Violence Homicide Victims
----------------	------------------------------------	--

You Can Also Experiment with the Extraction Method

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.



Stream



Lattice

Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

Now You Export Your Data and Do Analysis

Tabula My Files My Templates About Help Source Code Support Tabula on OpenCollective!

domestic-violence-report-2019.pdf Export Format: CSV Export Copy to Clipboard

Is the extracted data incorrect?
You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells
Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.
← Revise selection(s)

Choose Alternate Extraction Method
The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Latice** method instead.

Preview of Extracted Tabular Data

	New Hampshire Lethality Assessment Program (LAP) 2018-2019 Data	
	Law Enforcement Data	Crisis Center Data
		# of victims that
	# of victims # of victims that	
	# of screens # of screens	utilized follow-up
	screened in as spoke to	
	initiated completed	crisis center

Support Tabula on OpenCollective!

CSV Export Copy to Clipboard

ar Data



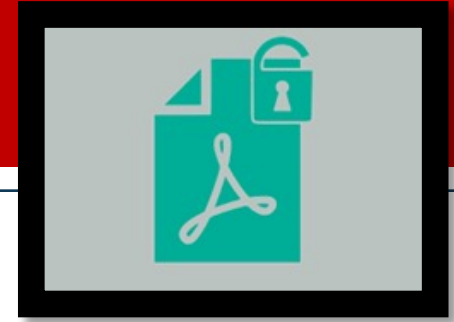
Let's Export The Data to A CSV



I Can Clean Up My Data or Delete Information I Don't Want

	A	B	C	D	E	F	G	H	I	J	K	L
36		2015	14 7 2	4 1 50% 14	4	100%						
37			# Domestic Violence -									
38		2016	14 Partner	3 0 29% 79	1	0%						
39		2017	11 5 2	3 0 45% 18	1	100%						
40		2018	19 9 5	3 1 47% 47	5	100%						
41			1									
42		2019	# Do2m8 e	3 0 43% 32	9	44%						
43	Totals		170 87 46	34 7 51% 2	45	77%						
44	CHART 3: COMPARISON OF MURDER SUICIDES 2018-2019											
45	Year		Total # ofH	Total # ofM	# of Domes	# of Non-Domestic	Violence Murder/Suicide Victims					
46		2018	19	5	5	0						
47		2019	28	9	4	5						
48	CHART 5: HOMICIDE CAUSE OF DEATH 2018-2019											
49	Cause of Death		Domestic V	Non-Domestic Violence	Homicide Victims							
50	Firearm		10	15								
51	Hand Gun		9	15								
52	Long Gun		1	0								
53	Cut/Stab		4	4								
54	Blunt Impact		3	5								
55	Strangulation		2	0								
56	Other		2	2								

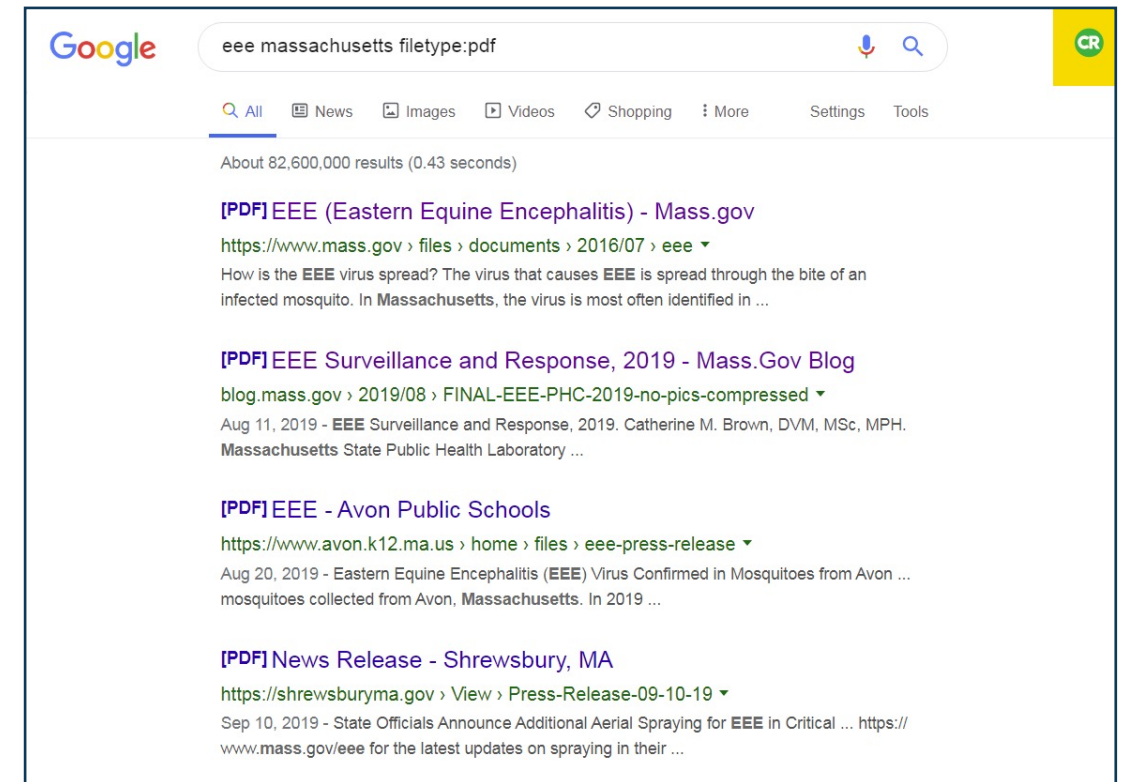
Export Your Data



- Select EXPORT
- Export to a CSV
- Open your spreadsheet
- Do a four corners check? (how many records do you have?)
 - What does your top row say?
 - What is your bottom record?

Search the web for EEE data for Massachusetts

- Use filetype: pdf and mass.gov in your search so we are sure we are getting government reports
- Click on the second link and find the table within the document
- Some things to notice:
- Sometimes government releases data in PowerPoint presentations!
- How old is the presentation?



The Link You Should Get

EEE data



Converting Data

- Open **Tabula** and convert the tabular data to Excel
- Open your Excel spreadsheet

- Do you see any story ledes?
- Try to write one out and dictate it to the class