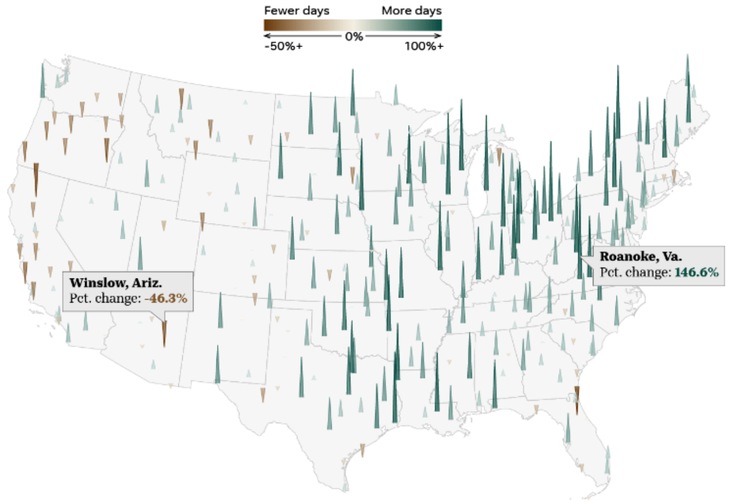
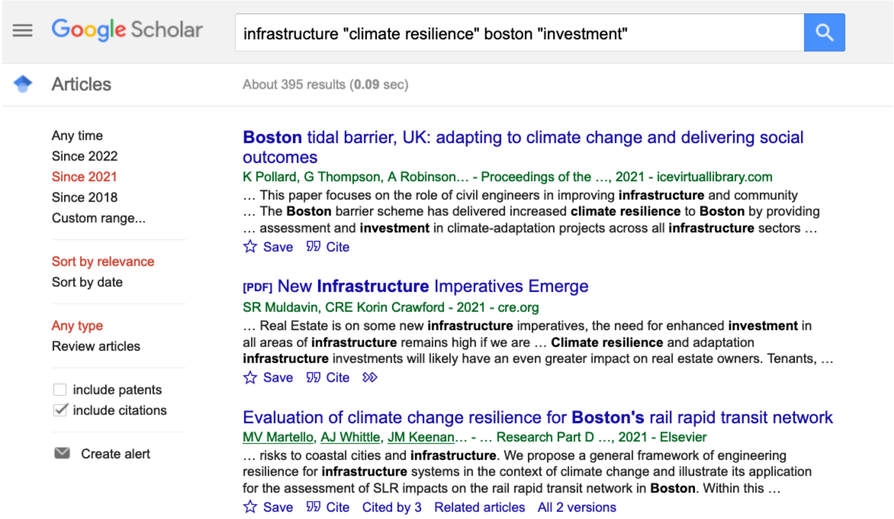
Some data journalism tips

Aleszu Bajak
Senior data reporter, USA TODAY
aleszu.com
Twitter: @aleszubajak
Email: abajak@usatoday.com

Slides to accompany this tipsheet

https://docs.google.com/document/d/1kCy-3eOlxHy2igCrkswoZBOR_muKJY_5DlRqkCLMqIU/edit?usp=sharing





To Whom It May Concern:

Pursuant to the Ohio Open Records Law, §149.43 et seq., I hereby request:

- a **spreadsheet, table or other document** listing the most current list of providers enrolled to provide COVID-19 vaccines in Ohio **including provider name, id, date enrolled, location, and all non patient-identifying fields**.

...

I request these records be provided in an electronic format that can be imported into standard database software. Examples of such formats include an Excel .xls or .xlsx file, an Access .mdb or .accdb file, a text-based delimited file such as .csv or tab-delimited .txt, a .dbf file or an SQL dump readable by standard open-source database software. (**A PDF file would not comply with this request** because PDF files are not readable by database software.)

	A	B
1	**date**	**title**
27	9/5/2020	"This is a sickness that cannot be allowed to continue. Please report any si
28	9/9/2020	Chris Rufo on Glenn Beck: With Biden, Critical Race Theory may return to
29	9/16/2020	Against Wokeness: Conservatives must understand the threat posed by cri
30	9/18/2020	Defending American Values, and History
31	9/18/2020	You're Racist? Then No More Federal Funding, DOE Calls Princeton's Bluf
32	10/4/2020	The Truth About Critical Race Theory: Trump is right. Training sessions for
33	10/5/20	Neo-Marxist CRT destroying America
34	10/17/20	White supremacy meaning (not a column just major mainstream coverage)
35	10/26/20	How Critical Race Theory is Dividing America
36	12/7/20	**Critical race theory, the new intolerance (42 page report)**
37	1/12/21	CRT infiltrates government, CLASSROOMS
38	1/25/21	Biden embraced lunatic CRT
39	2/2/21	We need to get parents, in particular, to focus on CRT that will be forced or
40	2/11/21	CRT conflicts with Christian worldview
41	2/15/21	Private Schools Aren't Safe From The Woke Mob

1 **FIND**

- Ideation/Inspiration
- Research/Interviews/FOIAs
- Literature/Coverage review
- Collect/Structure

2 ANALYZE

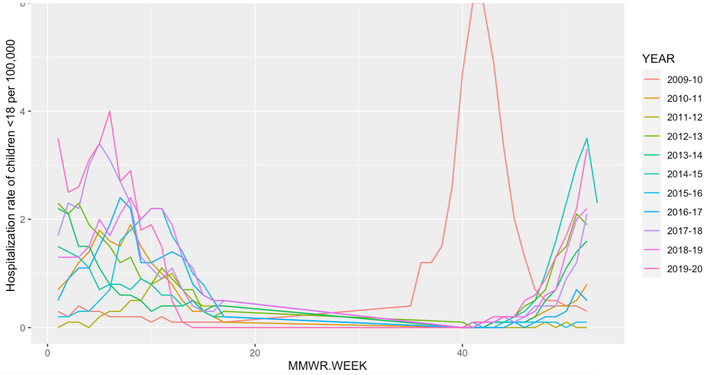
- Exploratory data analysis
- Recipe
- Merge

3 BULLETPROOF

- Dummy check
- Correlation does not equal causation

4 PRESENT

- Translate
- Nerdbox




```
# ***********************************
# Bring in Hopkins data
# ***********************************

# https://github.com/CSSEGISandData/COVID-19
# Download Zip file of Github repo
# Unzip and copy COVID-19-master to working dir
# Combine all files
# Filter for just 12/15/2021
```

Narrate over time
Start big and zoom in
Start small and zoom out
Highlight contrasts
Explore the intersections
Dissect the factors
Profile the outliers

Novelty
Outlier
Archetype
Trend
Debunking
Forecast

AL	AM	AN	AO
Population_	RUCC_2013	Description	doses_per_black_resi
5176	9	Nonmetro - Com	921.25
3558	9	Nonmetro - Com	104.5882353
10266	7	Nonmetro - Urba	121.2307692
5269	9	Nonmetro - Com	726.6666667
25857	7	Nonmetro - Urba	15.79910213
7259	9	Nonmetro - Com	45.9137931
10438	9	Nonmetro - Com	53.89705882
4552	9	Nonmetro - Com	23.71875
5896	9	Nonmetro - Com	78.76
9687	6	Nonmetro - Urba	26.28735632
10866	6	Nonmetro - Urba	36.61764706
6018	9	Nonmetro - Com	82.05882353
20428	7	Nonmetro -
12441	7	Nonmetro -
9726	7	Nonmetro -
9439	8	Nonmetro -
10995	8	Nonmetro -
14553	6	Nonmetro -

Doses per Black resident

1 FIND
- Ideation/Inspiration
- Research/Interviews/FOIAs
- Literature/Coverage review
- Collect/Structure

2 **ANALYZE**
- Exploratory data analysis
- Recipe
- Merge

3 BULLETPROOF
- Dummy check
- Correlation does not equal causation

4 PRESENT
- Translate
- Nerdbox

Hi Aleszu,

Thanks for reaching out and I'm happy to speak. But before we do, can you clarify the exact dates that your graph below shows? It seems like you're saying August 2020 – April 2021, is that right?

If that is the case, I think it's pretty important to include the Delta era in a graph like this, which I see you did in the linked USA Today piece.

Do you have a version that includes up to October 2021?

Thanks

> Hi Dr. Connor,
>
> I came across your name on a study on RSV in middle-aged and older adults – and I wondered if you might be able to help me locate pediatric (17 and under) hospitalization rates (per 100,000 children) for RSV by U.S. state for the past few seasons?
>
> I'm trying to compare those against Covid and influenza rates, which I've already found at HHS and CDC's FluServe-NET.
>
> I'm basically trying to fill out the following chart, but with RSV data. The chart shows hospitalization rates per 100,000 children for Covid vs. two flu seasons. I'd like to add an RSV line that would be most appropriate.

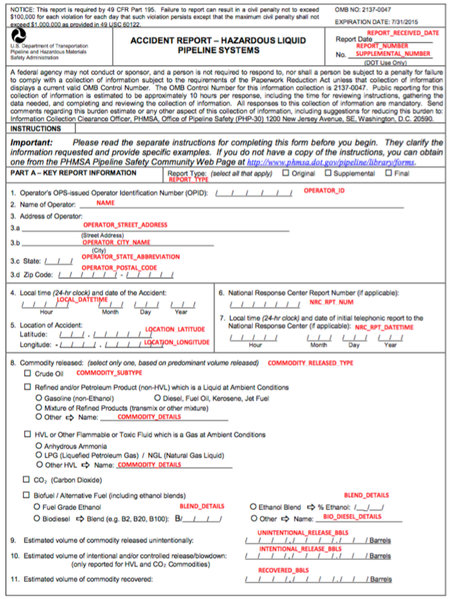
1 FIND
- Ideation/Inspiration
- Research/Interviews/FOIAs
- Literature/Coverage review
- Collect/Structure

2 ANALYZE
- Exploratory data analysis
- Recipe
- Merge

3 **BULLETPROOF**
- Dummy check
- Correlation does not equal causation

4 PRESENT
- Translate
- Nerdbox

LOCAL_DATETIME	IYEAR	LOCATION_L	LOCATION	NRC_RPT_N	NRC_RPT_DATETIME	COMMODITY	RECOVERED	NET_LOSS
1/26/15 9:38	2015	40.366817	-80.584594	1106602	1/26/15 11:38	HVL OR OTH	0	30565
1/10/15 20:45	2015	28.885881	-96.433027	1105356	1/10/15 22:01	HVL OR OTH	0	27123
3/12/13 17:46	2013	29.633617	-90.1741	1040890	3/12/13 23:55	HVL OR OTH	0	23702
8/12/13 23:10	2013	41.697667	-90.098627	1056922	8/13/13 0:17	HVL OR OTH	0	18400
7/29/13 0:00	2013	48.524251	-102.856912	1061615	9/30/13 0:16	CRUDE OIL	5856	14744
				146	2/25/10 7:00	HVL OR OTH	0	13718
				851	9/8/11 8:36	HVL OR OTH	0	12836
				38	8/13/13 16:02	HVL OR OTH	0	11405
				108	6/18/15 9:30	HVL OR OTH	0	8000
				71	6/4/11 9:31	CRUDE OIL	4383	7846
				62	9/10/16 19:16	CRUDE OIL	1560	7040



NOTICE: This report is required by 49 CFR Part 195. Failure to report can result in a civil penalty not to exceed $100,000 for each violation for each day that such violation persists except that the maximum civil penalty shall not exceed $1,000,000 as provided in 49 USC 60122.

OMB NO: 2137-0047
EXPIRATION DATE: 7/31/2015

U.S. Department of Transportation
Pipeline and Hazardous Materials Safety Administration

ACCIDENT REPORT – HAZARDOUS LIQUID PIPELINE SYSTEMS

Report Date: REPORT_RECEIVED_DATE
No. REPORT_NUMBER
SUPPLEMENTAL_NUMBER
(DOT Use Only)

A federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with a collection of information subject to the requirements of the Paperwork Reduction Act unless that collection of information displays a current valid OMB Control Number. The OMB Control Number for this information collection is 2137-0047. Public reporting for this collection of information is estimated to be approximately 10 hours per response, including the time for reviewing instructions, gathering the data needed, and completing and reviewing the collection of information. All responses to this collection of information are mandatory. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: Information Collection Clearance Officer, PHMSA, Office of Pipeline Safety (PHP-30) 1200 New Jersey Avenue, SE, Washington, D.C. 20590.

INSTRUCTIONS

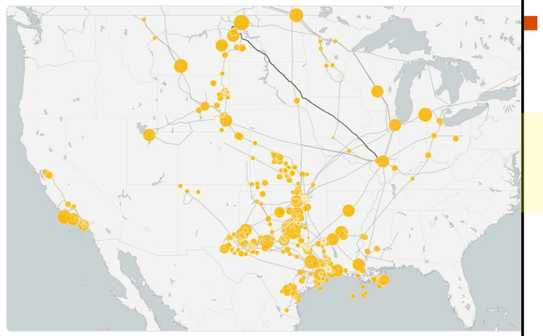
Important: Please read the separate instructions for completing this form before you begin. They clarify the information requested and provide specific examples. If you do not have a copy of the instructions, you can obtain one from the PHMSA Pipeline Safety Community Web Page at http://www.phmsa.dot.gov/pipeline/library/forms.

PART A – KEY REPORT INFORMATION

Report Type: *(select all that apply)* ☐ Original ☐ Supplemental ☐ Final
REPORT_TYPE

1. Operator's OPS-issued Operator Identification Number (OPID): OPERATOR_ID
2. Name of Operator: NAME
3. Address of Operator:
   3.a OPERATOR_STREET_ADDRESS (Street Address)
   3.b OPERATOR_CITY_NAME (City)
   3.c State: OPERATOR_STATE_ABBREVIATION
   3.d Zip Code: OPERATOR_POSTAL_CODE
4. Local time *(24-hr clock)* and date of the Accident: LOCAL_DATETIME (Hour / Month / Day / Year)
5. Location of Accident: Latitude: LOCATION_LATITUDE; Longitude: LOCATION_LONGITUDE
6. National Response Center Report Number (if applicable): NRC_RPT_NUM
7. Local time *(24-hr clock)* and date of initial telephonic report to the National Response Center (if applicable): NRC_RPT_DATETIME (Hour / Month / Day / Year)
8. Commodity released: *(select only one, based on predominant volume released)* COMMODITY_RELEASED_TYPE
   ☐ Crude Oil COMMODITY_SUBTYPE
   ☐ Refined and/or Petroleum Product (non-HVL) which is a Liquid at Ambient Conditions
   ○ Gasoline (non-Ethanol) ○ Diesel, Fuel Oil, Kerosene, Jet Fuel
   ○ Mixture of Refined Products (transmix or other mixture)
   ○ Other ⇨ Name: COMMODITY_DETAILS
   ☐ HVL or Other Flammable or Toxic Fluid which is a Gas at Ambient Conditions
   ○ Anhydrous Ammonia
   ○ LPG (Liquefied Petroleum Gas) / NGL (Natural Gas Liquid)
   ○ Other HVL ⇨ Name: COMMODITY_DETAILS
   ☐ CO_2 (Carbon Dioxide)
   ☐ Biofuel / Alternative Fuel (including ethanol blends) BLEND_DETAILS
   ○ Fuel Grade Ethanol BLEND_DETAILS ○ Ethanol Blend ⇨ % Ethanol:
   ○ Biodiesel ⇨ Blend (e.g. B2, B20, B100): B ○ Other ⇨ Name: BIO_DIESEL_DETAILS
9. Estimated volume of commodity released unintentionally: UNINTENTIONAL_RELEASE_BBLS / Barrels
10. Estimated volume of intentional and/or controlled release/blowdown: (only reported for HVL and CO_2 Commodities) INTENTIONAL_RELEASE_BBLS / Barrels
11. Estimated volume of commodity recovered: RECOVERED_BBLS / Barrels

Here's an interactive map of all pipeline spills since 2010. @undarkmag #noDAPL #standingrock undark.org/article/oil-pi



10:43 AM - 26 Jan 2017

COVID-19 in kids: Methodology

Weekly pediatric admissions were calculated by totaling each week's previous day pediatric admissions confirmed positive for COVID-19 reported by hospitals to HHS. Rates were calculated by dividing total weekly admissions by child population sourced from the 2019 American Community Survey. Cases and deaths per 100,000 by age group were sourced directly from the CDC. The child vaccination rate was calculated by subtracting the number of people age 18 or older fully vaccinated from the number of people age 12 or older fully vaccinated, then dividing by the child population.

Over 70,000 miles of crude oil pipelines — most of them underground — spread like vasculature beneath the earthen skin of the United States. Those conduits, **sometimes wide enough for a person to shimmy through**, transport around one billion gallons of oil every day from production fields to refineries, and they help to satisfy the nation's oil habit, which amounts to about 19.4 million barrels (or 814 million gallons) of oil consumed every day, on average. But with that much oil and that much pipe — **enough to circle the globe nearly three times** — accidents are inevitable. In fact, since 2010, there have been more than 1,300 crude oil spills in the United States, according to data collected by the Pipeline and Hazardous Materials Safety Administration, a regulatory arm of the U.S. Department of Transportation.

That's one crude oil spill every other day.

1 FIND
- Ideation/Inspiration
- Research/Interviews/FOIAs
- Literature/Coverage review
- Collect/Structure

2 ANALYZE
- Exploratory data analysis
- Recipe
- Merge

3 BULLETPROOF
- Dummy check
- Correlation does not equal causation

4 **PRESENT**
- Translate
- Nerdbox

Tools for data analysis

Excel

Google Sheets

python

RStudio

Google colab

Tools for data visualization

Datawrapper

RAWGraphs

CARTO

plotly

ArcGIS StoryMaps

infogr.am

Easy intros to coding

RStudio

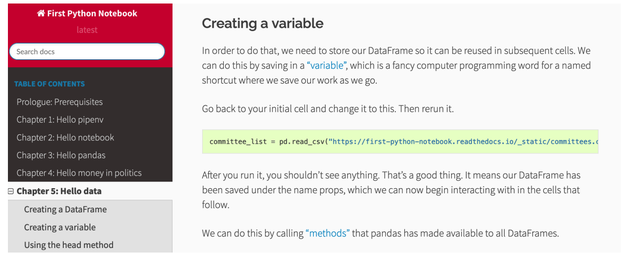
Introduction

USING R FOR DATA JOURNALISM

This course is designed to give you a sense of all the possibilities from programming in R.

https://learn.r-journalism.com/en/introduction/

python



https://www.firstpythonnotebook.org