INTERNATIONAL CIVIL UNREST

KALEB KOLB
PSDS Capstone
15 January 2021

INTERNATIONAL CIVIL UNREST

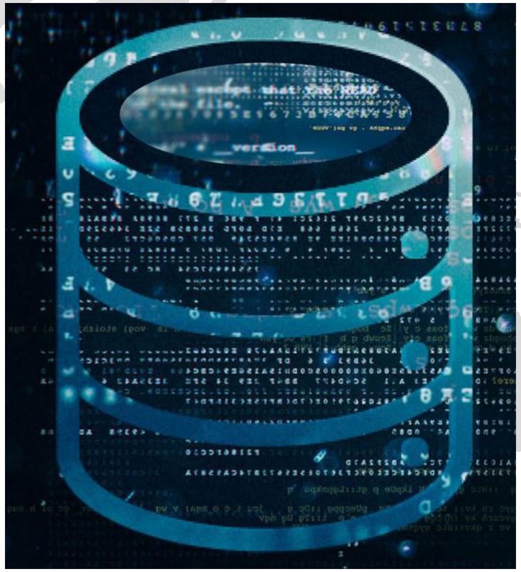
- Question to Answer
- Data Acquisition
- Data Carpentry
- Exploratory Analysis
- Modeling
- Conclusions
- Way Forward
- Lessons Learned

QUESTION TO ANSWER

- What Factors Contribute to Civil Unrest (Protests and Riots)?
  - Economic
  - Educational
  - Health
  - Infrastructure
  - Social
  - Other

DATA ACQUISITION

- Data Locations
  - World Bank Open Data (data.worldbank.org)
    - Indicators
  - Armed Conflict Location & Event Data (ACLED) Project (acleddata.com)
    - Protest and Riot Event Data
  - Both Tools Aggregate Data from Various Sources



DATA ACQUISITION

- World Bank Data
  - Thousands of Indicators to Choose From
  - Initially Picked 394 Factors
    - Economic (108), Education (61), Health (67), Infrastructure (46), Social Issues (34), Other (78)
  - Analyzed 92 Factors, Years 2010-2019, Worldwide
    - Economic (17), Education (15), Health (10), Infrastructure (19), Social Issues (11), Other (20)
  - Data is Pulled via API Call
    - http://api.worldbank.org/v2/country/USA/indicator/SI.DST.10TH.10?format=json
    - Single call per factor, per country, returns JSON string
      - 92 Factors x 216 Countries and Territories = 19,872 API Calls

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

DATA ACQUISITION

- World Bank Data


```
[{"page":1,"pages":2,"per_page":50,"total":61,"sourceid":"2","lastupdated":"2020-12-16"},[{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2020","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2019","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2018","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2017","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2016","value":30.4,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2015","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2014","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2013","value":30,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2012","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2011","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2010","value":29.3,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2009","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2008","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2007","value":30.4,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2006","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2005","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2004","value":30.1,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2003","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2002","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2001","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"2000","value":30.3,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1999","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1998","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1997","value":30.4,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1996","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1995","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1994","value":29.6,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1993","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1992","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1991","value":27.5,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1990","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1989","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1988","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1987","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1986","value":26.7,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1985","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1984","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1983","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1982","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1981","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1980","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1979","value":25.3,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1978","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1977","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1976","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1975","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1974","value":26,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1973","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1972","value":null,"unit":"","obs_status":"","decimal":1},{"indicator":{"id":"SI.DST.10TH.10","value":"Income share held by highest 10%"},"country":{"id":"US","value":"United
States"},"countryiso3code":"USA","date":"1971","value":null,"unit":"","obs_status":"","decimal":1}]]
```

DATA ACQUISITION

- ACLED Data
  - Conflict data primarily in Africa, Asia, and Latin America
  - Individual events are documented, primarily from news sources
    - Can be called via API; I exported to 10 CSV files
  - Used data for Protests and Riots from 2010 - 2019

ACLED

DATA ACQUISITION

- ACLED Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	data_id	iso	event_id_	event_id_	event_date	year	time_prec	event_typ	sub_even	actor1	assoc_act	inter1	actor2	assoc_actor_2	inter2
2	7263461	586	PAK5379	5379	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
3	7263430	586	PAK5383	5383	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
4	6059942	788	TUN97	97	31-Dec-10	2010	1	Protests	Protest wi	Police Forces of Tuni		1	Protesters (Tunisia)	UGTT: Tunisian G	6
5	6059916	788	TUN99	99	31-Dec-10	2010	1	Protests	Excessive	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
6	6059874	788	TUN89	89	31-Dec-10	2010	1	Protests	Protest wi	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
7	6059753	788	TUN88	88	31-Dec-10	2010	1	Protests	Protest wi	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
8	6059492	788	TUN98	98	31-Dec-10	2010	1	Protests	Peaceful p	Protesters	Journalist	6			0
9	6059387	788	TUN95	95	31-Dec-10	2010	1	Protests	Excessive	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
10	6059281	788	TUN96	96	31-Dec-10	2010	1	Protests	Protest wi	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
11	6059270	788	TUN93	93	31-Dec-10	2010	1	Protests	Excessive	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
12	6059177	788	TUN90	90	31-Dec-10	2010	1	Protests	Excessive	Police Forces of Tuni		1	Protesters (Tunisia)	Lawyers (Tunisia	6
13	5926514	50	BGD9735	9735	31-Dec-10	2010	1	Riots	Mob viole	Rioters (B	Bohoramp	5	Rioters (Bangladesh)	Bohorampur Cor	5
14	5590613	50	BGD9742	9742	31-Dec-10	2010	1	Riots	Mob viole	Rioters (B	Vigilante (	5	Civilians (Bangladesh)		7
15	5590611	50	BGD9740	9740	31-Dec-10	2010	1	Riots	Mob viole	Rioters (B	BCL: Bangl	5	Civilians (Banglades	BNP: Bangladesh	7
16	5590609	50	BGD9738	9738	31-Dec-10	2010	1	Riots	Mob viole	Rioters (Bangladesh)		5	Rioters (Bangladesh)		5
17	5590608	50	BGD9737	9737	31-Dec-10	2010	1	Riots	Mob viole	Rioters (Bangladesh)		5	Rioters (Bangladesh)		5
18	5533261	144	SRI1297	1297	31-Dec-10	2010	1	Riots	Mob viole	Rioters (Sr	UPFA: Uni	5	Civilians (Sri Lanka)	UNP: United Nat	7
19	5529186	586	PAK5407	5407	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
20	5529185	586	PAK5406	5406	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
21	5529184	586	PAK5405	5405	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
22	5529183	586	PAK5404	5404	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
23	5529182	586	PAK5403	5403	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
24	5529181	586	PAK5402	5402	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
25	5529180	586	PAK5401	5401	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
26	5529179	586	PAK5400	5400	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
27	5529177	586	PAK5399	5399	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0
28	5529176	586	PAK5398	5398	31-Dec-10	2010	1	Protests	Peaceful p	Protesters (Pakistan		6			0

DATA ACQUISITION

- Data Pull and Storage
  - Used nested loop to iterate through the indicators and countries and make API calls to World Bank dataset
  - Pulled ACLED data in from CSV files
  - Saved everything to PostgreSQL Database in 93 tables
  - ~30 seconds per country per category (Econ, Health, Social, etc…)
    - ~10 hours to populate database

```
YEM  Economic Data input Complete
--- 7320.4866988658905 seconds ---
ZMB  Economic Data input Complete
--- 7357.522523641586 seconds ---
ZWE  Economic Data input Complete
--- 7392.47868180275 seconds ---
Econ Tables Complete
--- 7392.479679107666 seconds ---
```

DATA ACQUISITION

- PostgreSQL – *unrestdatabase*
  - 93 Tables
    - acled_201X, countries, indicator_factors, indicator_201X,

acled_2010	econ_factors	education_2010	health_2010	infra_2010	other_2010	social_2010
acled_2011	economic_2010	education_2011	health_2011	infra_2011	other_2011	social_2011
acled_2012	economic_2011	education_2012	health_2012	infra_2012	other_2012	social_2012
acled_2013	economic_2012	education_2013	health_2013	infra_2013	other_2013	social_2013
acled_2014	economic_2013	education_2014	health_2014	infra_2014	other_2014	social_2014
acled_2015	economic_2014	education_2015	health_2015	infra_2015	other_2015	social_2015
acled_2016	economic_2015	education_2016	health_2016	infra_2016	other_2016	social_2016
acled_2017	economic_2016	education_2017	health_2017	infra_2017	other_2017	social_2017
acled_2018	economic_2017	education_2018	health_2018	infra_2018	other_2018	social_2018
acled_2019	economic_2018	education_2019	health_2019	infra_2019	other_2019	social_2019
countries	economic_2019	education_factors	health_factors	infra_factors	other_factors	social_factors

DATA ACQUISITION

- Table Examples
  - ***acled_2010***

Data Output

	index bigint	country text	Protest_Count_2010 double precision	Riot_Count_2010 double precision	Country_Code text	Region text
1	0	Algeria	8	10	DZA	Africa
2	1	Banglade...	576	1138	BGD	Asia
3	2	Benin	9	[null]	BEN	Africa
4	3	Botswana	[null]	1	BWA	Africa
5	4	Burkina F...	3	1	BFA	Africa
6	5	Burundi	3	3	BDI	Africa
7	6	Cambodia	180	22	KHM	Asia
8	7	Cameroon	4	[null]	CMR	Africa

economic_2010

Data Output

	COUNTRY.CODE character (3)	NY.ADJ.NNTY.KD.ZG character (255)	NY.ADJ.NNAT.GN.ZS character (255)	MS.MIL.TOTL.TF.ZS character (255)	IC.I cha
12	AUT	1.91194085083917 ...	8.25029688918482 ...	0.602306808419294 ...	Nor
13	AZE	-4.87695617954473 ...	41.4685931499104 ...	1.83944314753551 ...	Nor
14	BHS	-2.77749344782195 ...	16.9148704626987 ...	0.444347996817228 ...	21
15	BHR	15.1970291474788 ...	27.1105422746195 ...	2.72500549622126 ...	Nor
16	BGD	6.43845386061935 ...	27.8412500726495 ...	0.386768891999102 ...	Nor
17	BRB	None ...	-6.72454220398793 ...	0.399902974360319 ...	1.2
18	BLR	7.40595725496453 ...	13.3416621966833 ...	3.59383778814155 ...	Nor
19	BEL	1.42735909209834 ...	5.97905995316616 ...	0.731185867013627 ...	Nor

DATA CARPENTRY

- Data Initially looked to be in fairly good shape upon retrieval
- Fairly well-structured and when looking at it from a macro scale seemed complete enough to continue
- Didn’t think I would have to do much more to the data

- And then I started exploratory data analysis…

EXPLORATORY ANALYSIS

- Wanted to find indicators that correlated with # of Protests and # of Riots
- Knew a negative correlation is as good as a positive
- Looped through indicators and years to find correlations with absolute value > 0.6 (somewhat arbitrary)
  - Saved to master set to de-dupe

```
{'EN.POP.DNST',
 'IC.FRM.BRIB.ZS',
 'IC.TAX.GIFT.ZS',
 'SE.TER.CUAT.BA.ZS',
 'SH.DTH.1519',
 'SH.DTH.2024',
 'SI.DST.FRST.10',
 'SP.HOU.FEMA.ZS',
 'SP.POP.TOTL',
 'SP.RUR.TOTL',
 'SP.URB.TOTL'}
```

EXPLORATORY ANALYSIS

Data Output

	COUNTRY.CODE character (3)	IC.TAX.GIFT.ZS character (255)
1	AFG	None
2	ALB	None
3	DZA	None
4	ASM	None
5	AND	None
6	AGO	34.2
7	ATG	6.1
8	ARG	8.7
9	ARM	None
10	ABW	None

Data Output

	COUNTRY.CODE character (3)	IC.FRM.BRIB.ZS character (255)
57	DOM	12.3
58	ECU	None
59	EGY	15.2
60	SLV	4.2
61	GNQ	None
62	ERI	None
63	EST	None
64	SWZ	6.7
65	ETH	None
66	FRO	None

Data Output

	COUNTRY.CODE character (3)	SE.TER.CUAT.BA.ZS character (255)
69	FRA	17.4036998748779
70	PYF	None
71	GAB	None
72	GMB	None
73	GEO	32.8466300964355
74	DEU	25.5801200866699
75	GHA	None
76	GIB	None
77	GRC	20.1387195587158
78	GRL	None

Data Output

	COUNTRY.CODE character (3)	SP.HOU.FEMA.ZS character (255)
188	CHE	None
189	SYR	None
190	TJK	None
191	TZA	24.5
192	THA	None
193	TLS	17.5
194	TGO	None
195	TON	[null]
196	TTO	None
197	TUN	None

EXPLORATORY ANALYSIS

- Needed a way to fill in ‘None’ values
  - Pandas has function – Interpolate
- At this point data is in table by year, so back to carpentry

pandas.DataFrame.interpolate

`DataFrame.interpolate(method='linear', axis=0, limit=None, inplace=False, limit_direction=None, limit_area=None, downcast=None, **kwargs)`

Fill NaN values using an interpolation method.

DATA CARPENTRY

- Current table format was not conducive to interpolation
- Joins are expensive
- SQL → DF → SQL

interpolate verb

Save Word

in·ter·po·late | \ in-ˈtər-pə-ˌlāt \

interpolated; **interpolating**

Definition of *interpolate*

transitive verb

1 **a** : to alter or corrupt (something, such as a text) by inserting new or foreign matter

b : to insert (words) into a text or into a conversation

2 : to insert between other things or parts : INTERCALATE

3 : to estimate values of (data or a function) between two known values

DATA CARPENTRY

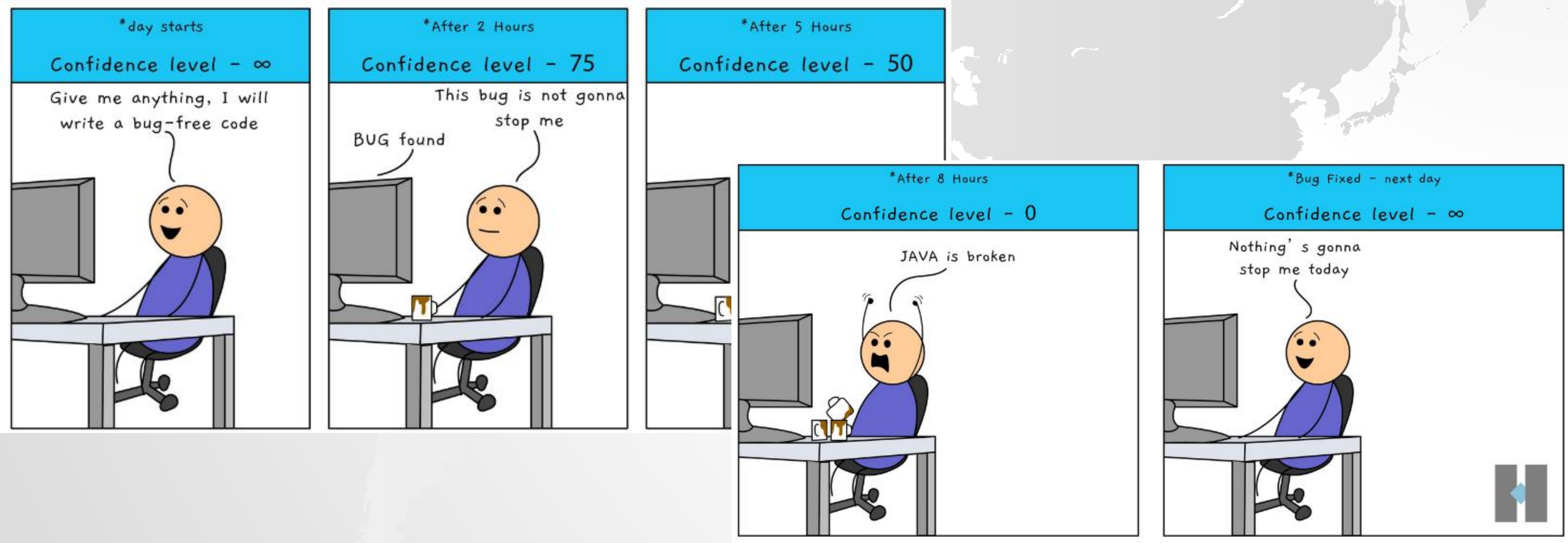
- Table Example
  - ***BGD_2010***

Data Output

	index text	Year_2010 double precision	Year_2011 double precision	Year_2012 double precision	Year_2013 double precision	Year_2014 double precision	Year_2015 double precision	Year_2016 double precision	Year_2017 double precision	Year_2018 double precision	Year_2019 double precision	Feature text	Country_Code text
1	NY.ADJ.NNTY.KD.ZG	6.43845386061935	5.42606862995126	6.25578337364061	5.71030918986794	4.32142926097241	6.0994909063875	6.66179382369818	6.21826101343281	8.03955937502239	8.15268494677524	NY.ADJ.N...	BGD
2	NY.ADJ.NNAT.GN.ZS	27.8412500726495	27.2799154961492	29.5146277053423	28.7838273328652	27.2900367255922	26.3524343824762	26.9036645305456	25.5015619495541	23.5432386476418	8.15268494677524	NY.ADJ.N...	BGD
3	MS.MIL.TOTL.TF.ZS	0.386768891999102	0.379674272295274	0.372751263495194	0.365988568832801	0.359383616840298	0.352978694464805	0.346868722865313	0.327626295749028	0.330056511547617	8.15268494677524	MS.MIL.T...	BGD
4	IC.FRM.BRIB.ZS	7.34568564347453	7.51025744772915	7.0395551449582	47.7	7.2443622329628	7.42907677964477	8.61476795652378	11.4176407399684	1.40629744280687	8.15268494677524	IC.FRM.B...	BGD
5	GC.DOD.TOTL.GD.ZS	14.30460239495	14.640840623163	13.7063590264212	33.4897306757689	14.1293408490853	14.5051748648247	16.8826671901822	22.5076551841877	2.48253837406612	8.15268494677524	GC.DOD....	BGD
6	GC.XPN.COMP.ZS	21.2635191464254	21.7714237985969	20.3731629078842	19.2794613515377	21.0143194652078	21.5812729500047	25.1505664238407	33.597669628407	3.55877930532537	8.15268494677524	GC.XPN....	BGD
7	per_allsp.cov_pop_tot	17.7709074941925	17.7955773915827	16.7949548494489	36.964281304431	17.1776918072917	17.3894769234506	40.991202882407	44.6876840726264	4.63502023658462	8.15268494677524	per_allsp....	BGD
8	SL.EMP.TOTL.SP.NE.ZS	56.5969009399414	13.8197309845685	13.2167467910137	54.6491012573242	13.3410641493756	13.1976808968964	56.0900001525879	55.7776985168457	5.71126116784387	8.15268494677524	SL.EMP.T...	BGD

EXPLORATORY ANALYSIS

- After 2500 lines of code, I'm back



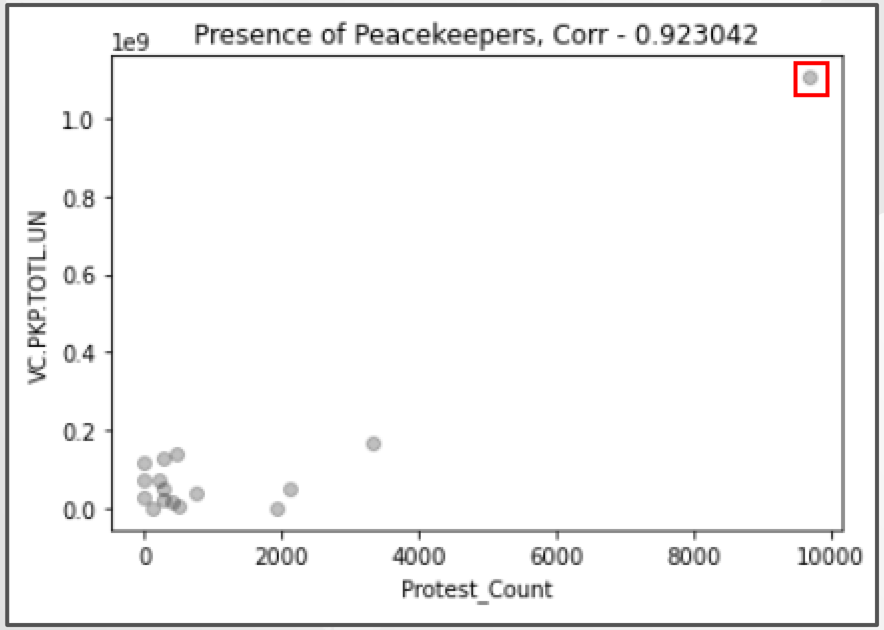
EXPLORATORY ANALYSIS

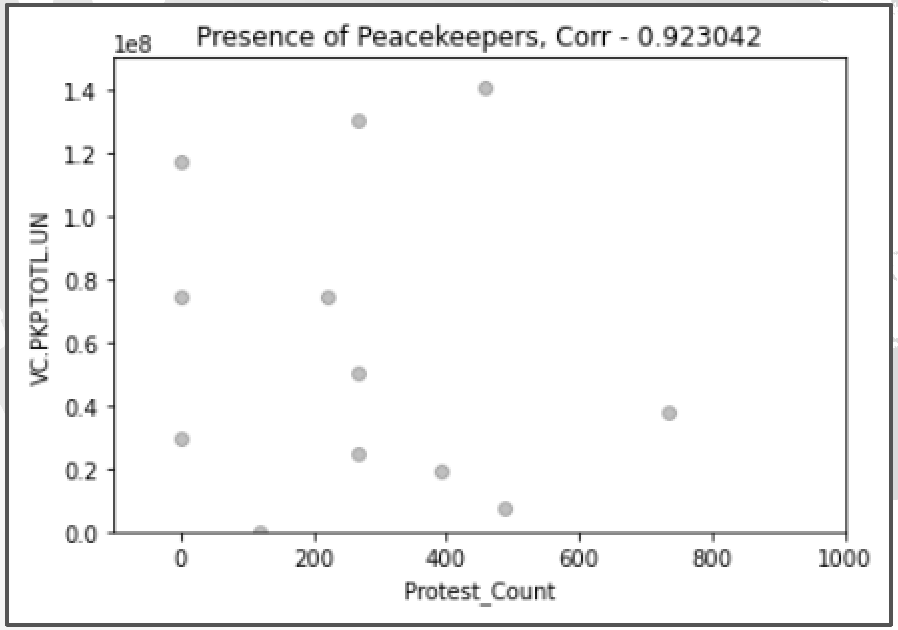
- Looped through indicators and years to find correlations with absolute value > 0.75 (still somewhat arbitrary)
  - Saved to master set to de-dupe

EN.POP.SLUM.UR.ZS - Population living in slums (% of urban population)	*SP.DYN.CDRT.IN - Death rate, crude (per 1,000 people)*
SH.DTH.1014 - Number of deaths ages 10-14 years	*SP.POP.TOTL - Population, total*
SH.DTH.1519 - Number of deaths ages 15-19 years	*SP.RUR.TOTL - Rural population*
SH.DTH.2024 - Number of deaths ages 20-24 years	*SP.URB.TOTL - Urban population*
SH.XPD.CHEX.PC.CD - Current health expenditure per capita (current USD)	*VC.PKP.TOTL.UN - Presence of peace keepers*

EXPLORATORY ANALYSIS

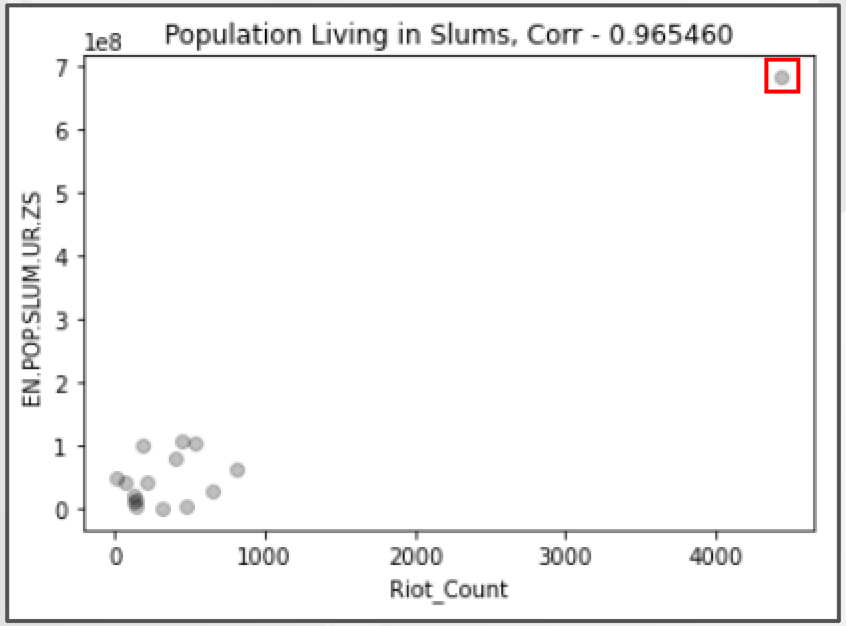
- Looped through indicators and years to find correlations with absolute value > 0.75 (still somewhat arbitrary)
  - Saved to master set to de-dupe

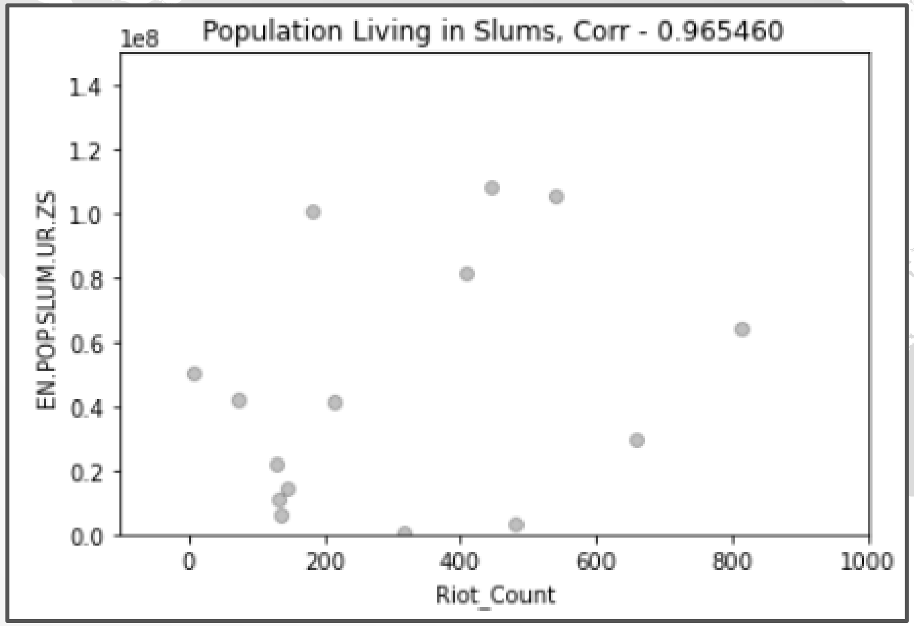




EXPLORATORY ANALYSIS

- Looped through indicators and years to find correlations with absolute value > 0.75 (still somewhat arbitrary)
  - Saved to master set to de-dupe





STATISTICAL MODELING

- Regression – Ordinary Least Squares (OLS)
  - Dep. Variable: Protest_Count
  - Ind. Variables: 10 Indicators with correlation Value > 0.85
  - statsmodel.api package

R-squared (uncentered):	0.944
Adj. R-squared (uncentered):	0.872
F-statistic:	13.15
Prob (F-statistic):	0.00131
Log-Likelihood:	-125.91
AIC:	269.8
BIC:	276.8

	coef	std err	t	P>|t|	[0.025	0.975]
SP.RUR.TOTL	-0.0001	0.000	-0.822	0.438	-0.000	0.000
SP.URB.TOTL	-9.779e-06	8.28e-06	-1.181	0.276	-2.94e-05	9.79e-06
SH.XPD.CHEX.PC.CD	2.0665	0.867	2.383	0.049	0.016	4.117
VC.PKP.TOTL.UN	0.0003	0.000	0.922	0.387	-0.000	0.001
SH.DTH.1519	0.0477	0.090	0.532	0.611	-0.164	0.259
SH.DTH.1014	0.0201	0.078	0.256	0.805	-0.165	0.205
EN.POP.SLUM.UR.ZS	-1.319e-05	9.61e-05	-0.137	0.895	-0.000	0.000
SP.DYN.CDRT.IN	-16.0350	52.840	-0.303	0.770	-140.982	108.912
SP.POP.TOTL	-0.0001	0.000	-0.882	0.407	-0.000	0.000
SH.DTH.2024	-0.0882	0.085	-1.035	0.335	-0.290	0.113

WAY FORWARD

- Incorporate More Statistical Modeling
  - Additional Regression, ANOVA, MANOVA
- Machine Learning
- Find way to more reliably interpolate
- Find better way to compare countries
  - Region, Quality Index, Population, etc…
- Build Dashboard
- Sentiment Analysis – Triggers

LESSONS LEARNED

- Start with a smaller data set
- Be flexible
- Put everything in a database
- API vs. CSV?
- Play to your Strengths