Data Commons

datacommons.org

Context

There is a lot of data ...

From census (US, India, UK, ...), RBI, USGS, BLS, BEA, UDISE, World Bank, ...

This data is essential

For science, journalism, policy and our users

Using this data is extremely painful

Repeated expensive data wrangling

Analogy with Satellite Imagery and Google Earth/Maps





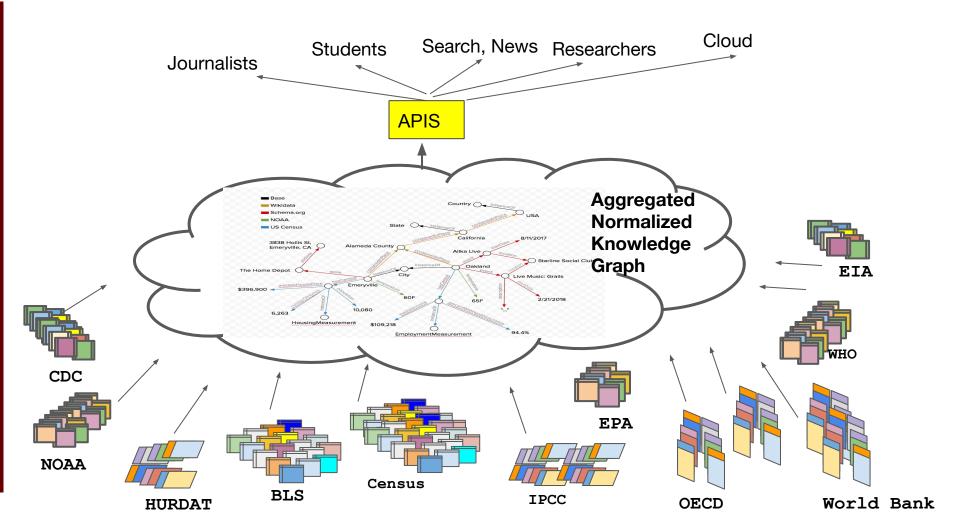
Our goal for data

	sibea Data - Tools -	News - Research - Res	ources - About - Help				
	National Data GDP & Personal Income	National Data Industry Data Fixed Assets GDP-by-indu		rational Da I Transact			
		International Data Direct Investment & MNEs	Regional Data GDP & Personal Income				
1	What is the Interactive	Data Application?	Census			Q Search	
	EA's interactive data application is the		BROWSE BY TOPIC	EXPLO	RE DATA	A LIBRARY SURVEYS/ PROGRAMS	INFORMATION FOR
		Center for on Statistics				munity Survey	
 2005 Date 2003 Date 	ta Products ta Products		America	100	JIII	intunity Survey	
	ta Products		About the Survey	D	ata Ta	ables & Tools	
	ta Products		Respond to the Survey				
• <u>1991-199</u>	36 Data Products		News & Updates	Da	ta Profi	iles Selector	
			Data Data Tables & Tools			les consist of four tables (Social, Economic, Housing, Demo	
2010 0.4	TA DEODUISTC	Colored The	Data Tables & Tools			view of a particular geography. We provide an easy to use D or the most popular geographies: state, county and place. Ou	
	1 15 1	(A)		da	ta.censu	us.gov, provides additional geographies for this table type an	id many more.
	8-48			Na	rrative I	Profiles	
	WMO	UNEP		Th	e only pl	lace to find the current Narrative Profiles is right here on Am	erican Community Survey
	IPCC PI		100000			larrative Profiles contain much of the same information four s a text-based report with plenty of colorful graphs and chart	
			IPCC Secretari	81			
	IPCC B	COLORD BY		ss		Public-Use Data Files a	nd
	Executive C	ommittee		ath	+		nu ACS
-	-		-	_		Documentation	atile
Working Group I	Working Group II	Working Group III	Task Force on	Data		The National Center for Health Statistics (NCHS) is	On This Page
The Physical	Impacts,	Mitigation	National			pleased to offer downloadable public-use data files	NHANES
Science Basis	Adaptation,	of	Greenhous	ed	+	through the Centers for Disease Control and Prevention's (CDC) FTP file server. Users of this servic	
and the second	and	Climate Change	Gas	lle		have access to data sets, documentation, and	e National Hea Surveys most
	Vulnerability		Inventories			guestionnaires from NCHS surveys and data collection systems. Downloading instructions are available in	NVSS p
TSU	TSU	TSU	TSU	t		"readme" files.	
		A CONTRACTOR		ersity		Public-use data files are prepared and disseminated	NSFG
	Authors, Contribu	tors, keviewers		n for rd Socia	ы	to provide access to the full scope of the data. This allows researchers to manipulate the data in a forma	NHIS
PELR file	📬 (10.5 MB)		Research			appropriate for their analyses. NCHS makes every	NIS
	ll file 💷 (11.2 MB)		SPAC	E Program		effort to release data collected through its surveys and data systems in a timely manner.	LSOA
	file 街 (43.2 MB)		Vital	Statistics			
	public-use setup files (3 SAS ASCII Read-In code, f		Online			Users of NCHS public-use data files must comply with data use restrictions to ensure that the information	SLAITS
	ote that the ASCII file is not					will be used solely for statistical analysis or reporting	Data Linkage
NHES: 2016 PFI Public-Use File Codebook 🔂 (1.4 MB)			Deleted Citer			purposes.	Compressed
Adult Training and Education (ATES)			Related Sites				File
			Data Linkage			National Health and	SPACE Progr
			NCHS Data			Nutrition Examination	
			Visualization			Survey (NHANES)	
			Galler	y.		Survey (minites)	

CA counties most at food risk from climate change?

From search for datasets, download, clean, normalize, join ...

to Just ask in Natural Language



What we have built: Code

Open source:

- GCP infrastructure for creating, storing, serving, KG
- Visualization tools

In Google's Search Stack: Integration with Google search

What we have built: Data

People & places, ...

- **Demographics**: Census (US, India, ...), Eurostat, ...
- Economics: BLS, BEA, WorldBank, ...
- Health: CDC, DEA, WHO, ICD, ...
- Food, Crime, Education, Elections, Trade, ...

Climate Change & Sustainability:

- Climate: IPCC, EPA, HURDAT, NOAA
- Energy: EIA, NREL, ...
- Water, Agriculture: USGS, WRIS (India), USDA, FAO
- Emissions: EPA, EIA, ..

3.5+ Billion time series
1.7 trillion triples
2.4 M places
100+ k variables
3x search KG
5x FRED

What we have built: Data

Biomed

- Biochemistry: PharmGKG, ChEMBL, Drugs@FDA ...
- **Protein:** UniProt, MINT, HUPO, Tissue Atlas
- Sequence: clinVar, ENCODE, dbSNP, GTEx, NCBI Gene...
- Species: Uniprot controlled vocab of species
- **Omics:** HMDB, Human1D, Virtual Metabolic Human
- Misc: SIDER, CDC Wonder, ICD-10, Disease Ontology...

Principal Investigator is Samantha Piekos

Work started at Stanford, now at The Institute for Systems Biology

44M Gene variants 3.4M Bio specimens 1.95M compounds

Application: In Google Search

NL interface in Google search

Energy use per capita in India

CO2 emissions in Sweden

Number of unemployed in California

Population growth rate in Germany

Fertility rate in bangladesh

Number of women phds in Mountain View

Number of poor hispanic women in Santa Clara

1:14 M (C) (C)	▼
Google	8
Q number of hispanic wo	omen in sa 🌷
All News Images Videos	Maps Shopping
Santa Clara County / Woman / His	spanic / Population
242,406 (2019)	
300,000	
200,000	Google 😡
100,000	Q fertility rate in bangladesh
0 2012 2014	
San Mateo County Alame 93,239 (2019) 183,31	
🕥 Explore more	Bangladesh / Fertility rate
Sources include: United States Census Burea	2.01 births per woman (2019)
Show more 🗸	
People also ask	
What percentage of Santa Clara Co	2
* ビ	Q 1960 1970 1980 1990 2000 2010 Pakistan India 3.45 (2019) 2.20 (2019)
	Static (2019) 2220 (2019)

Sustainability Data Commons

Climate Change: two facets

The Physical World: materials science, chemistry, grid, ...

The Information World

Understanding the impact of the changing climate to prepare

Reducing waste

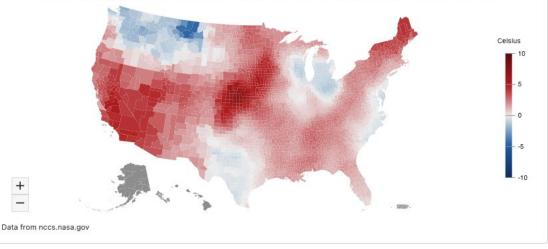
Carbon accounting

Climate Change: Temperature

Climate change is not as simple as 1.5 vs 2 degree.

Climate deltas vary widely across places

Max Temperature (Difference Relative To 2006), Based on RCP 2.6 (2050-06)



Top Places

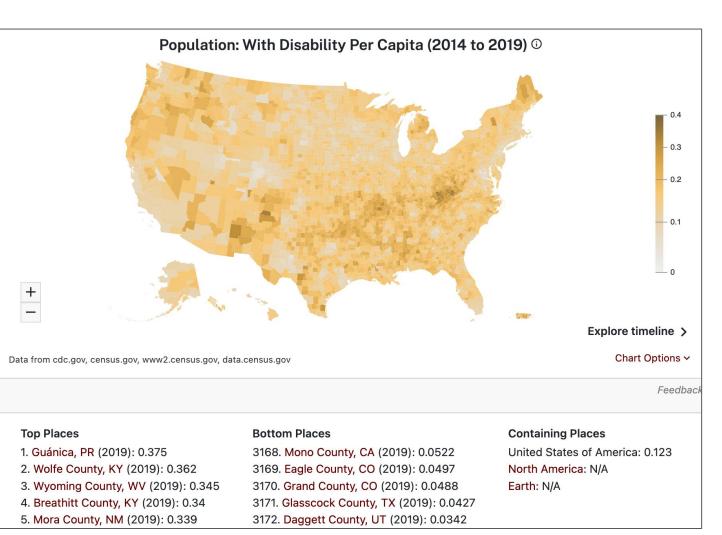
1. Phillips County, KS (2050-06): 7.22 Celsius 2. Norton County, KS (2050-06): 7.17 Celsius 3. Harlan County, NE (2050-06): 7.12 Celsius 4. Phelps County, NE (2050-06): 7.08 Celsius 5. Furnas County, NE (2050-06): 6.99 Celsius

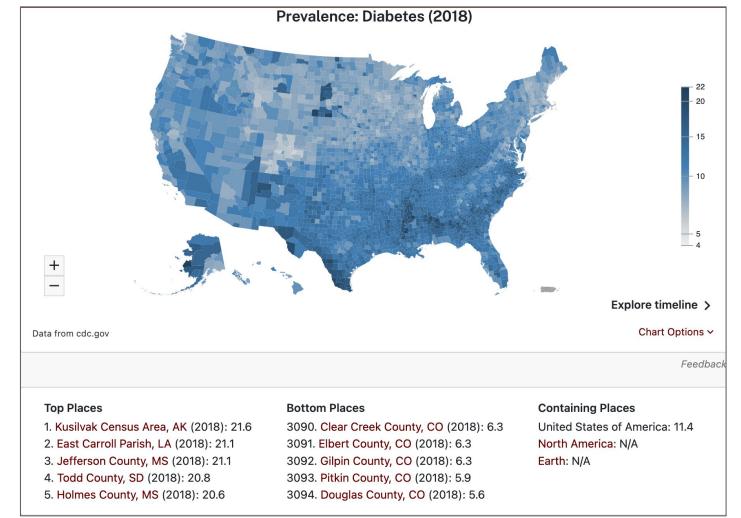
Bottom Places

3059. Petroleum County, MT (2050-06): -3.8 Celsius 3060. Chouteau County, MT (2050-06): -4.13 Celsius 3061. Blaine County, MT (2050-06): -4.27 Celsius 3062. Phillips County, MT (2050-06): -4.43 Celsius 3063. Fergus County, MT (2050-06): -4.48 Celsius

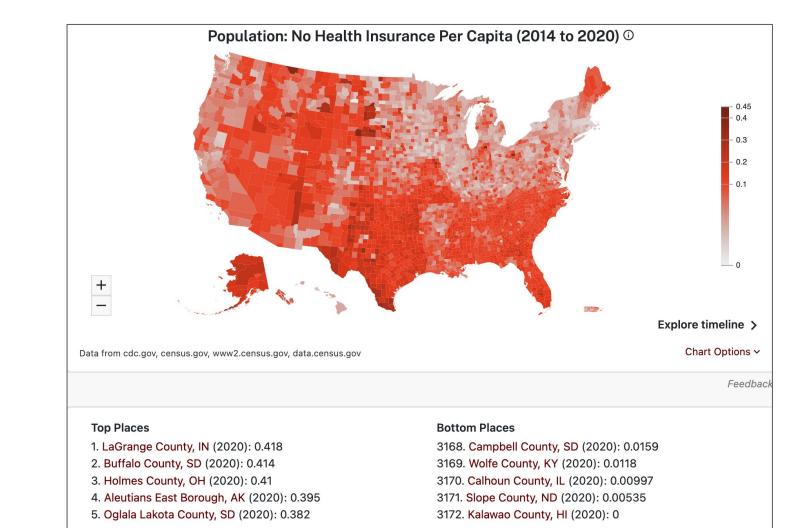
There are many inequities







Percentage of population with Diabetes



Without Health Insurance

Climate change will worsen these ...

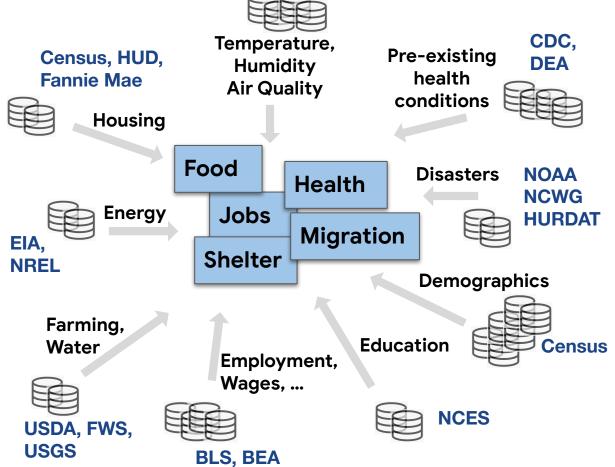
We have to prepare

To prepare, we have to know who is going to be most affected: 10, 20, 30 years from now

Messy Data Ecosystems

IPCC, NASA, NOAA, EPA

We need data not just about climate, but about data about health, food, farming, water, employment, ...



The Sustainability Data Commons Mission

- `Organize and make easily accessible' all the sustainability data for everyone
 - researchers, policy folks, administrators, students ... all of our users!
- Sustainability Data Commons is the base data layer for many applications
 - Different interfaces, data augmentations, ML tools and more
 - From many different companies, universities, ...
- This 'base data layer' is open and built together with other players from industry, academia and government (like html, schema.org, etc.)

Many topics already in and many more coming

Climate, Water, Agriculture, Air quality

Fuels, Energy Production

Natural Disasters (Storms, fires, floods, ...)

Mortality causes, Poverty, Social Benefits

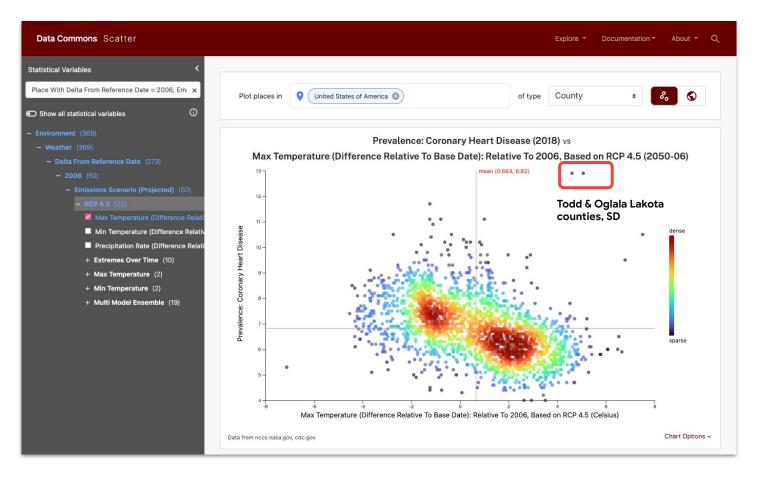
Education (K-12 and college)

Employment & labor markets, Businesses

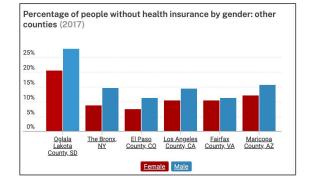
Housing, Commute, Insurance, Disabilities, Immigration, Crime

Application: The Resiliency Dashboard

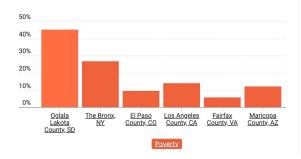
Health: Cardiac conditions --- who is in most danger from temperature rise?

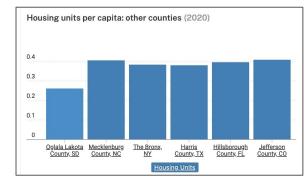


Oglala Lakota County --- other vulnerability factors



Poverty rate: other counties (2020)



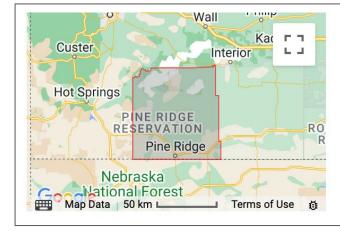


High rate of uninsured

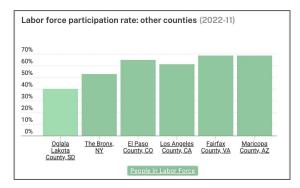
High Level of Poverty

Fewer homes per capita

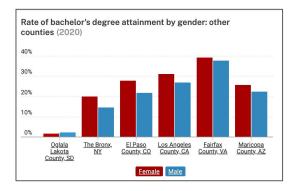
Oglala Lakota County --- the bigger picture



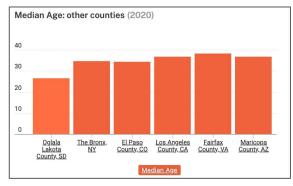
Rankings (in)	South Dakota	United States
Largest Population	<u>15 of 67</u>	2186 of 3108
Highest Median Income	<u>64 of 67</u>	3110 of 3187
Highest Median Age	<u>62 of 67</u>	3168 of 3187
Highest Unemployment Rate	<u>2 of 66</u>	<u>193 of 3185</u>
Data from <u>census.gov</u> , <u>fbi.gov</u> , <u>bls.gov</u>		



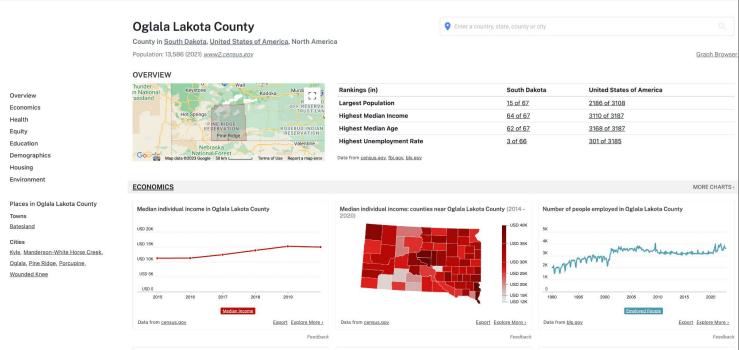
Low labor force participation



Low levels of college edu.



Very young!

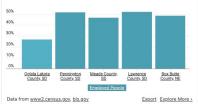


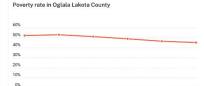
2015

2016

Data from www2.census.gov, census.gov

Percentage of people employed: counties near Oglala Lakota County (2022-12)



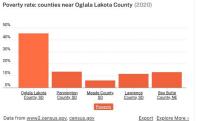


2017

2018

2019

Export Explore More >

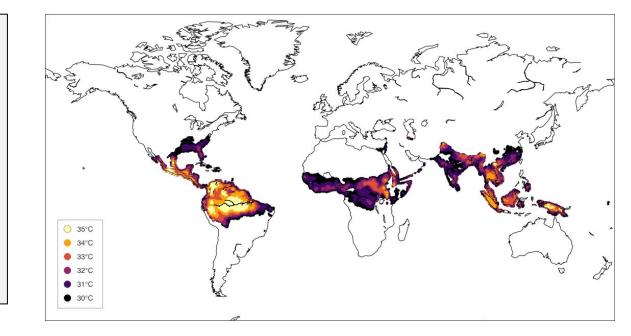


Not just the US ...

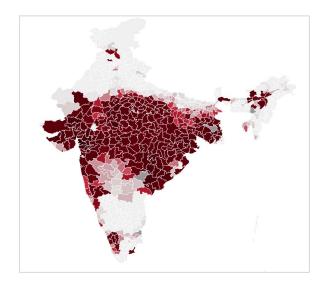
Wet Bulb Temperature

Unlike raw temperature, no IPCC data

Need models of humidity, intra-day variation in humidity, etc. to project max wet bulb temperature.



Focussing on India ...



Regions of India that will experience days with wet bulb temp > 34 deg

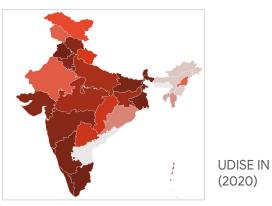
(color indicates number of days) (RCP 4.5, NASA_Mean_ACCESS1-0)

Who will be affected most?

Homeless Population



Census IN (2011)



Percentage of schools with functioning toilets

Glaciers feeding the Indus, Ganga & Brahmaputra

Celsius

Max Temperature (Difference Relative To Base Date): Relative To 2015, Based on RCP 8.5, SSP 5 (2050-06)

Obert Ortione .

Dashboards focus attention



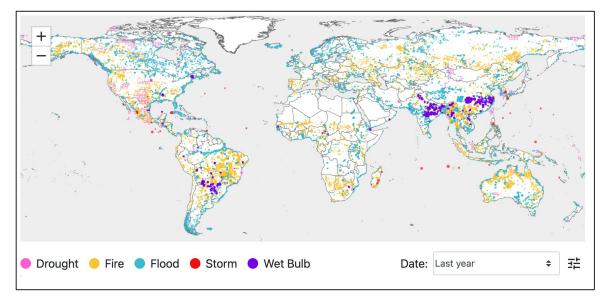
Coming Soon: Google-Stanford Resilience Platform

The platform that will inform about who is most at what kind of risk

A set of dashboards --- But more than that --- a platform on top of which many applications can be built

Part of the Google-Stanford

Data Commons partnership



Education

Data is the platform

- Berkeley DS100
- Being used data science/ML courses (in CS & Econ) at Berkeley, Harvard & MIT
 - Case Study: Prevalence of Obesity in 500 US Cities

Obesity is well known to correlate with health factors such as high blood pressure, but is also known to correlate with economic factors s low-income, unemployment, etc [1][2]. The Center for Disease Control (CDC) provides prevalence percentages on health conditions such a obesity, high blood pressure, and high cholesterol for approximately 500 major cities in the US (e.g. <u>San Francisco, New York</u>, and <u>Austin</u>) Meanwhile, the US Bureau of Labor Statistics provides <u>unemployment rates</u> while the US Census provides <u>poverty rates</u> for most cities ac the United States.

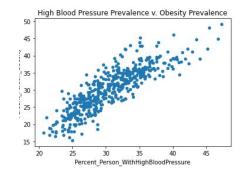
Even though these statistics come from different datasets across different government agencies with different storage formats, Data Commons surfaces each of these in a single, uniform knowledge graph. In fact, you can see this in the <u>browser</u> by looking at the *provenar* column. Let's use the data in Data Commons to create a linear regression model that incorporates variables:

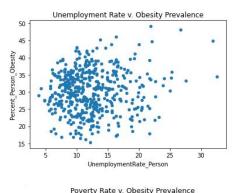
- · Prevalence of high blood pressure
- Unemployment rate
- · Percent of population living with income below the poverty line

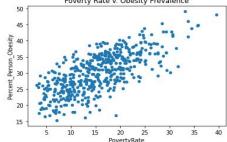
to predict the prevalence of obesity in the 500 cities that the CDC provides data for. One thing you may note is that the US Census also pr employment statistics (you can see this by navigating to the "employment" and "employmentStatus" sections for <u>San Francisco</u> and obse the different provenances). Our choice of using statistics from the Bureau of Labor Statistics is purely demonstrative, but it would be inte to see if similar results can be reproduced using US Census employment statistics.

[] !pip install datacommons_pandas --upgrade --quiet

famout determine mender of de







Data Commons in Core ML Curriculum (MIT)

- MIT's large undergraduate course: Intro to Machine Learning (6.036)
- Course offered every semester
- > 500 students (~400 Undergraduates; 100+ Masters students)
- Core, Required Course for ML/AI tracks
- All integration work was done by course TA's with minimal help from the core DC team

Data Literacy

Data Literacy with Data Commons

- Introductory, few prerequisites
- Pedagogical Goals:
 - Use Real Data
 - Be Interactive
 - Easy to Adapt
 - Focus on storytelling
 - Assume little programming and statistics background
- Target Audience:
 - Intro courses in data analysis/literacy/science
 - Specialized skills development, e.g. decision making for public policy, healthcare etc.
 - PreReqs: **no** programming, **descriptive** statistics

Data Science with Real Data

- Intermediate / Advanced
- Pedagogical Goals:
 - Use Real Data
 - Be Interactive
 - Easy to Adapt

Target Audience:

- Foundational courses in data analysis/literacy/science
- Specialized skills development, e.g. analytical tools for public policy, healthcare etc.
- PreReqs: programming, statistics

Biomedical Data Commons

Biomed DC: the data

Biochemistry Data

antibodyABCD, ChEMBL, Drugs@FDA,
 FDA - Pharmacologic Class, PharmGKB

Protein Data

 The Human Protein Tissue Atlas, HUPO -Proteomics Standard Initiative, Molecular INteraction Database (MINT), UniProt

Sequencing Data

 dbSNP, ENCODE, GTEx, NCBI - Assembly, NCBI - Gene, clinVar

<u> Disease / Healthcare Data</u>

 CDC 500 Cities / CDC Places, CDC Diabetes Surveillance System, CDC Wonder Mortality, ICD-10 Codes, Medical Subject Headings (MeSH), Side Effect Resource (SIDER)

Omics Data (coming soon)

 The Human Metabolome Database (HMDB), Human1D, Virtual Metabolic Human

Sample Application: Personalized medicine

Purpose: Clinician support for interpreting patient genetic information by putting into the context of disease and treatment.

English: For a genetic variant, identify the clinical significance and the associated disease. Then identify the disease description, symptoms, drug treatments, and drug dosages.



```
query = '''
SELECT ?gv dcid ?disease name
WHERE {{
?gv typeOf GeneticVariant .
?gv name "{0}" .
?gv dcid ?gv dcid .
?gv diseaseName ?disease name
}}
'''.format(GENETIC_VARIANT)
result = dc.query(query)[0] # get disease name
gv_dcid, disease_name = result['?gv_dcid'], result['?disease_name'].lower()
# get variant clinical significance
sigLink = list(dc.get_property_values([gv_dcid],\
                                      'clinicalSignificance').values())[0]
sig = list(dc.get property values(sigLink, 'name').values())[0][0]
query = '''
SELECT ?disease dcid ?description
WHERE {{
?disease typeOf Disease .
?disease name "{0}" .
?disease dcid ?disease dcid .
?disease description ?description
}}
'''.format(disease_name)
result = dc.query(query)[0] # get disease description
disease dcid, des = result['?disease dcid'], result['?description']
```

GENETIC VARIANT = 'rs33930977' # declare genetic variant

```
# identify symptoms associated with the disease
dis symp dcids = list(dc.get property values([disease dcid], \
                      'diseaseOntologyID', out=False, limit=1000).values())[0]
dis_symp_dcids = [dcid for dcid in dis_symp_dcids if "cooccurrence" in dcid]
symptoms dcids = list(dc.get property values(dis symp dcids,
                                             'medicalSubjectHeadingID').values())
symptoms dcids = list(itertools.chain.from iterable(symptoms dcids))
symptoms_names = list(dc.get_property_values(symptoms_dcids, 'name').values())
symptoms names = list(itertools.chain.from iterable(symptoms names))
# identify drug treatments
disease treat dcids = list(dc.get property values([disease dcid],)
                              'diseaseID', out=False, limit=1000).values())[0]
disease treat dcids = [dcid for dcid in disease treat dcids if "CTD " in dcid]
treatment_dcids = list(dc.get_property_values(disease_treat_dcids, \
                                              'compoundID').values())
treatment dcids = list(itertools.chain.from iterable(treatment dcids))
treatment_names = list(dc.get_property_values(treatment_dcids,)
                                      'commonName').values())
# get drug dosage forms
dosage_form = list(dc.get_property_values(treatment_dcids,\
                                          'dosageForm').values())
info = [sig, disease name, des, symptoms names, treatment names, dosage form]
for item in info: # print desired info
 print(item)
```

Uses data from 6 sources: ClinVar, DiseaseOntology, MedicalSubjectHeadings, UCSF SPOKE, ChEMBL, Drugs@FDA

Beyond A Single DataCommons.org

Limits of Google's datacommons.org

There is a lot of data that cannot come into Google's datacommons.org

Open data with non-commercial use licenses

Data govts not willing to share with Google

Open data behind walled gardens: D&B, Bloomberg, ...

Private: Patient records, sales, inventory

Bringing Data Commons benefits to this data

Many Data Commons, from different organizations

Different topics, some behind firewalls, some behind paywalls, ...

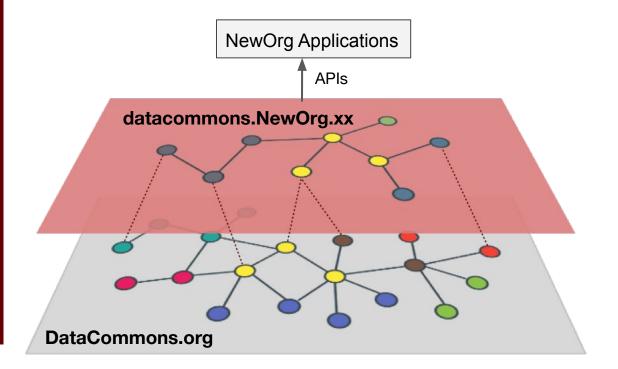
Like the Web

All have the same schema and same API

Same tools work on all the Data Commons

Why would someone put up a data commons?

- More value can be derived from their data



An overlay, on top of which both the overlay and base Data Commons can be accessed with the same, single API - overlaid data could be private or semi-public or ...

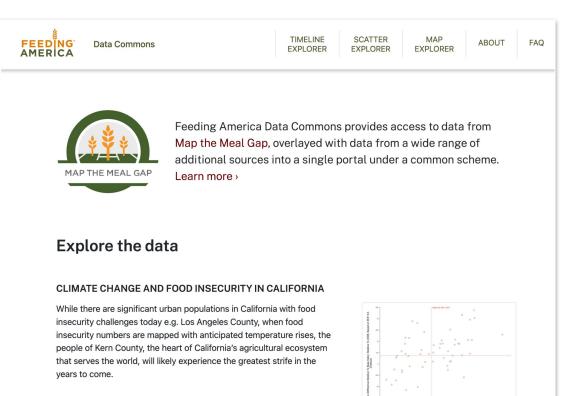
Feeding America and the Meal Gap Index



Food Insecurity among Overall (all ages) Population in the United States Before COVID-19

Click or touch on the map, or use the search bar to start exploring							
Search							
State, County, or Food Bank							-
Demographic		Мар Туре		Year			
Overall (all ages)	-	County	District	2020	2019	2018	2017

Case Study: Feeding America Data Commons



Max projected summer temperatures (RCP 4.5)

CA counties most at food risk from rising temperatures

From

From

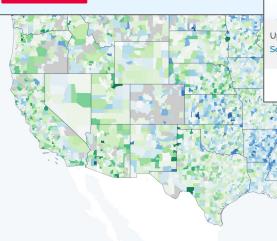
FEEDING Scatter AMERICA Statistical Variables % Plot places in California 🔯 of type County Search or explore below Show all statistical variables (\mathbf{i}) Max Temperature Relative To 2006 RCP 4 5 Difference Relative To Base Date vs + Agriculture (136) **Count Person Food Insecure Per Capita** DataCommons.org + Climate (198) Siskiyou 3.5 -+ Energy (10) Humboldt 2 Temperature Relative To 2006 RCP 4 5 Difference Relative Base Date (Celsius) + Crime (123) Trinity 3 -Del Norte 0 0 + Economics (20) 2.5 -+ Education (1738) 0 ^o Imperial 0 0 + Military Service (570) 0 0 0 0 2 -0 00 0 00 0 + Uncategorized (8) 1.5 -0 - [Feeding America] (9) 0 0 0 **FeedingAmerica** Annual_FoodBudgetShortfall Count_Child_FoodInsecure Santa Clara 0.5 -Count_Child_FoodInsecure_EligibleForManyChild 0 0 Count_Child_FoodInsecure_IneligibleForChildNut -0 0 Count_Person_FoodInsecure Count_Person_FoodInsecure_EligibleForAllFeder -0.5 -Count_Person_FoodInsecure_EligibleForSomeFe 0 -1 -Max Count_Person_FoodInsecure_IneligibleForFedera San Mateo 0 Mean_MealCost_Person_FoodSecure -1.5 -0.06 0.07 0.08 0.00 0.1 0 11 0.12 0 13 0 14 0.15 0.16 0.17 0.18 0.19 Count Person Food Insecure Per Capita

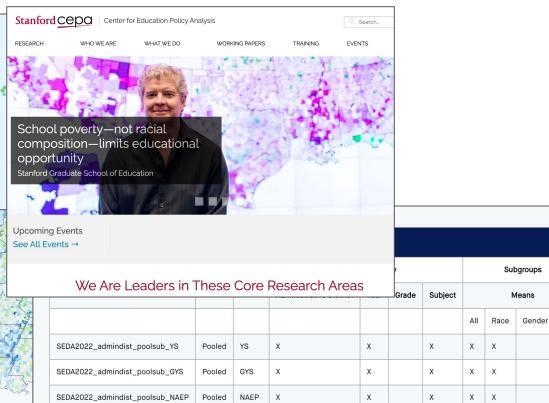
Case Study: Stanford CEPA

The 2009-2018 Educational Opportunity Explorer

View charts and maps that show how counties, school districts, and schools score on 3 key measures of educational opportunity. You can also filter by demographics, explore opportunity gaps, export PDF reports, and more.

GO TO THE EXPLORER





Download

Stata

Stata

Stata

CSV

CSV

CSV

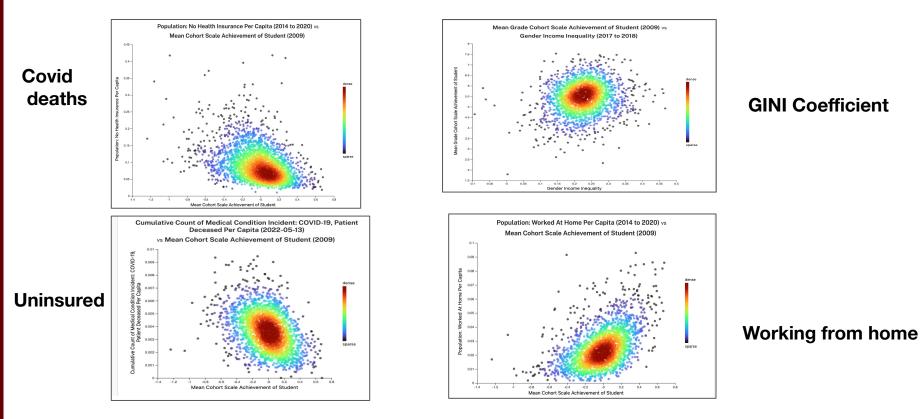
ECD

Х

Х

Х

How does this correlate with ...

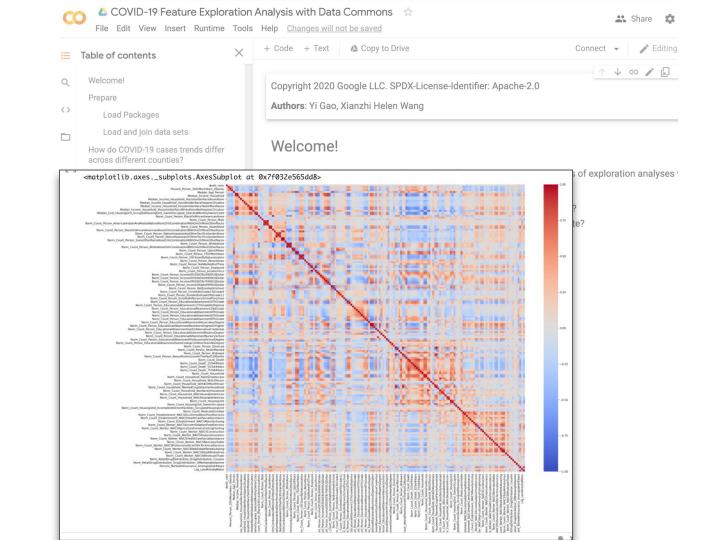


What if anyone could discover this in 10 min? Makes the CEPA data so much more useful!

Covid

Rate: Race Household size

Morbidity: Per Capita prescriptions



India Data Commons



EXPLORE T DOCUMENTATION T



India Data Commons

India Data Commons provides India-level data aggregated from a wide range of sources into a single portal under a common schema. India Data Commons is powered by Data Commons.



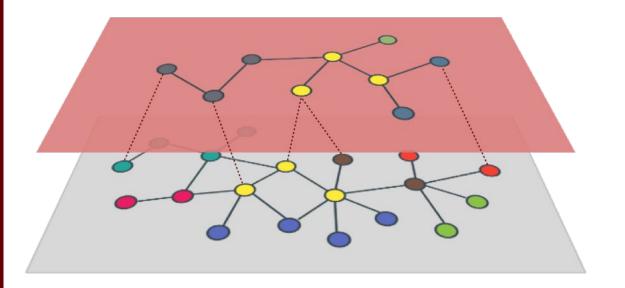
Learn more

Thanks

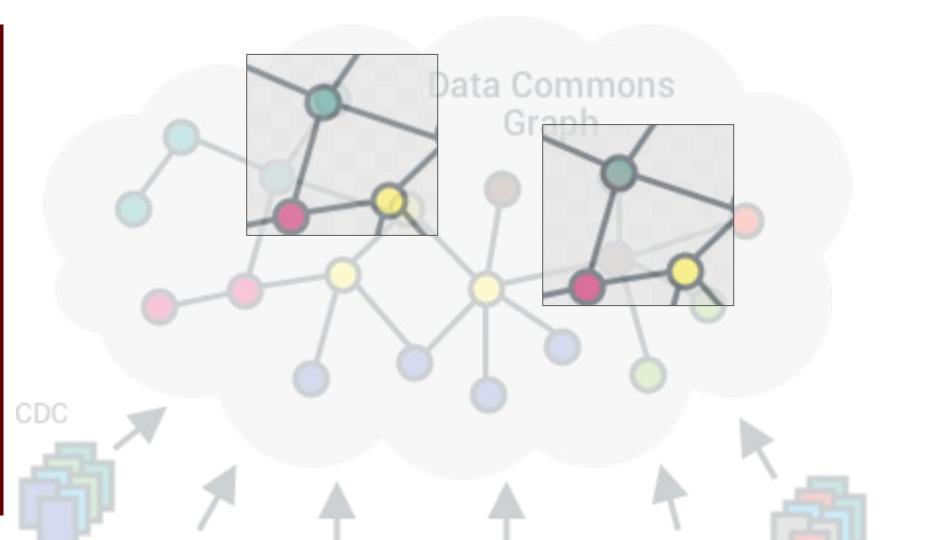


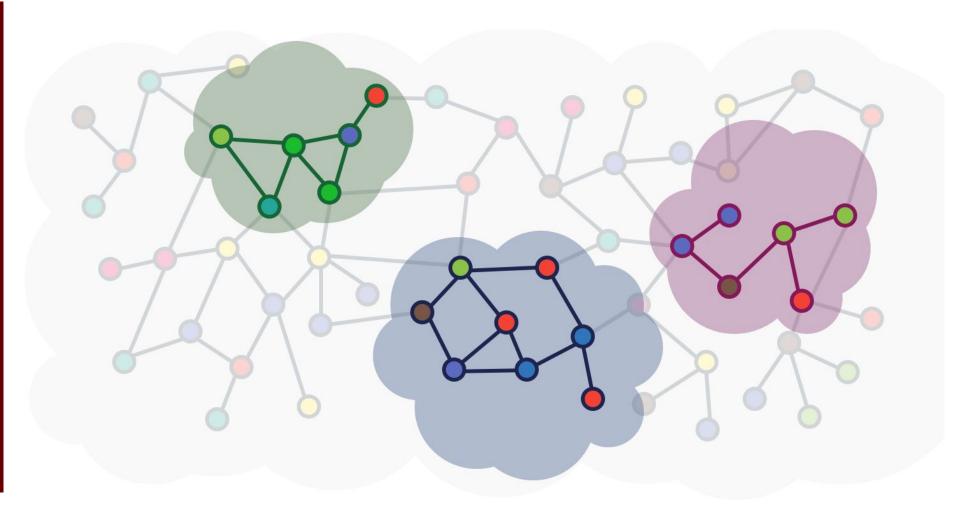
Many Data Commons, one schema, one API

- different topics, different ACLs, free vs paid, ...

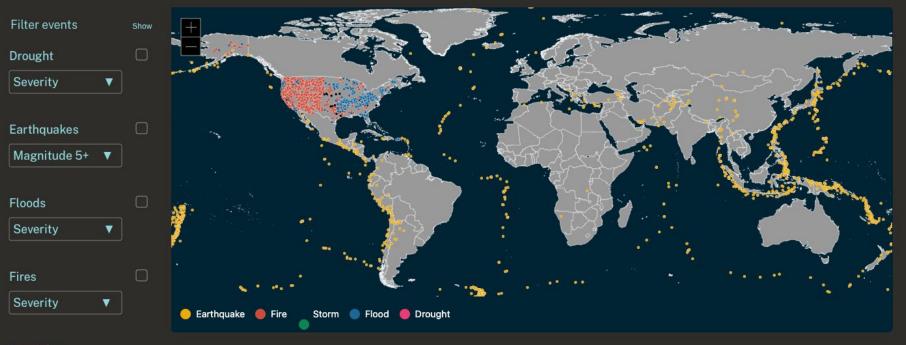


An overlay, on top of which both the overlay and base Data Commons can be accessed with the same, single API - overlaid data could be private or semi-public or ...





Stanford University Natural Disasters and Severe Weather Dashboard



Event dates

2022 🔻

Earthquakes in 2022



Fires in 2022



Storms in 2022



India Data Commons



EXPLORE T DOCUMENTATION T



India Data Commons

India Data Commons provides India-level data aggregated from a wide range of sources into a single portal under a common schema. India Data Commons is powered by Data Commons.



Learn more

Glaciers feeding the Indus, Ganga & Brahmaputra

Celsius

Max Temperature (Difference Relative To Base Date): Relative To 2015, Based on RCP 8.5, SSP 5 (2050-06)

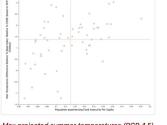
Obert Ortione .

Feeding America Data Commons

vears to come.



that serves the world, will likely experience the greatest strife in the

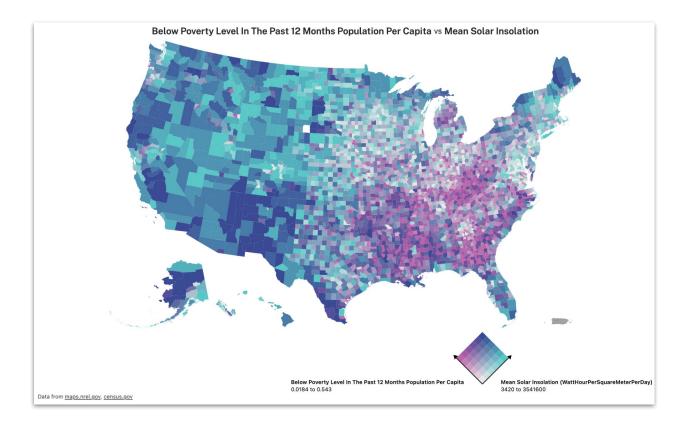


Max projected summer temperatures (RCP 4.5)

CA counties most at food risk from rising temperatures

FEEDING Scatter				Explore Documentation	A
Statistical Variables	Plot places in	California 😣	of type County	÷ %	¢
Search or explore below					
O Show all statistical variables	Max Temp	perature Relative To 2	006 RCP 4 5 Differend	e Relative To Base Date vs	5
Agriculture (136)		Count Pers	on Food Insecure Per C	Capita	
Climate (198)				Cielineur	
Energy (10)	۲		Humboldt	Siskiyou	
Crime (123)	ative		c	° Trinity	
Economics (20)	Rela	0 0	Del No	orte	
Education (1738)	0 2.5 –	0	0	0.	
Military Service (570)	2 - Differ	0 0	0 0	0 0	m
Uncategorized (8)	s) 12D	0	0 00 00		
[Feeding America] (9)	- 1.5 -		0 0 0	0	
Annual_FoodBudgetShortfall	000 H	0			
Count_Child_FoodInsecure	o 20 e Dat	0 0	0		
Count_Child_FoodInsecure_EligibleForManyChild	C - 5.0 Base	anta Clara		0	
Count_Child_FoodInsecure_IneligibleForChildNut	Lo ⁻ Relat	0 0			
Count_Person_FoodInsecure	ere F	0			
Count_Person_FoodInsecure_EligibleForAllFeder	Max Temperature Relative To 2006 RCP 4.5 Difference Relative To 2006 RCP 4.5 Difference Relative To				
Count_Person_FoodInsecure_EligibleForSomeFe	ше Настана Настана	0			
Count_Person_FoodInsecure_IneligibleForFedera	o S	an Mateo 🔗			
Mean_MealCost_Person_FoodSecure	-1.5		0.11 0.12 0.13 0.14	0.15 0.16 0.17 0.18 0.19	_

On a positive note: Where can renewable investment have maximal impact --- Solar Energy Potential vs. Poverty



Thank you!

A lesson from the past ...

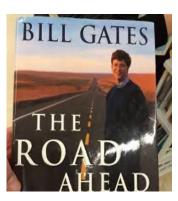
In 1993, a number of players wanted to 'own' the net

- Microsoft was planning to spend \$2B on MSN network
- AOL, the leader, was giving away 50m accounts
- IBM invested heavily in Prodigy
- Impressive tools & content partnerships for these networks
- Myhrvold's 10 reasons for why 'this web thing won't work
 - identity, security, commerce, ...









The Web

- Very amateurish
 - Mostly students in universities, researchers, enthusiasts
 - No authoring tool, no commerce, no security, ...
 - But, by end of '95, game over
 - May 26, '95, 'Internet Tidal Wave' memo
 - December '95, AOL, MSFT kill proprietary services

How did *that* happen?



The Web actually had far more investment!

- Web recruited millions of people contributing content
 - Even the billions spent by AOL/MSFT could not compete with this
- Everybody was putting up `interesting' pages and convincing 10 others to get on the web to see their page
 - Almost no barrier to entry

•

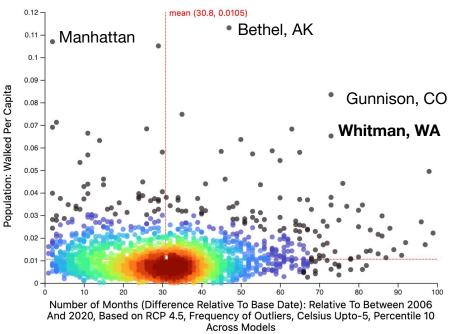
- Permissionless innovation --- no biz dev deals required
- Gradual investment/learning curves

Back in the US, so many 'vulnerability vectors'

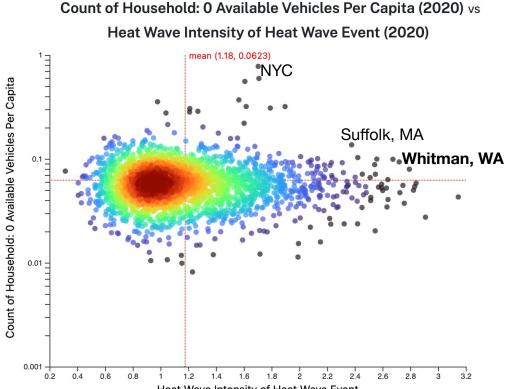
What about people who walk to work?

Population: Walked Per Capita (2020) vs

Number of Months (Difference Relative To Base Date): Relative To Bet 2006 And 2020, Based on RCP 4.5, Frequency of Outliers, Celsius Up Percentile 10 Across Models (2050)

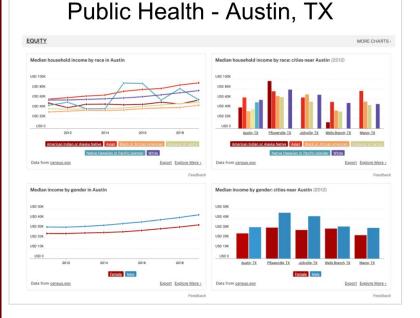


What about people without cars?

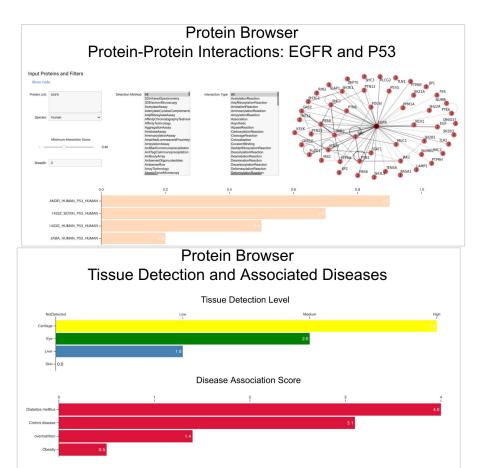


Heat Wave Intensity of Heat Wave Event

Biomedical Data Commons site (coming Q1 2023)



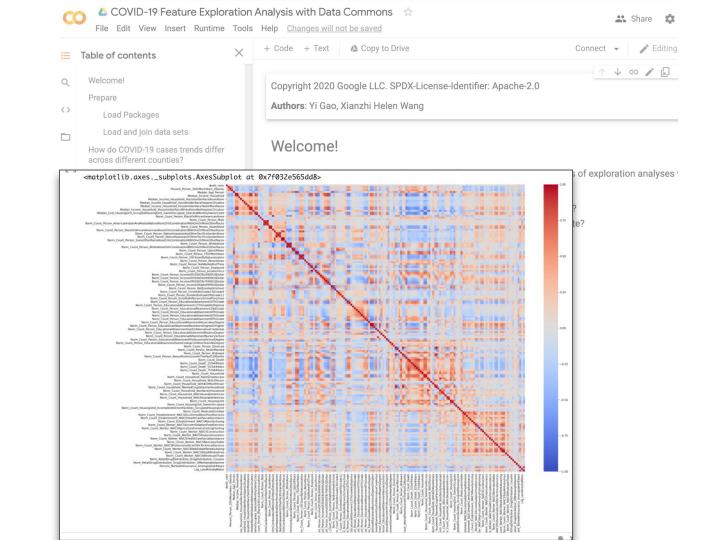
Pages for each disease, protein, place, ...



Covid

Rate: Race Household size

Morbidity: Per Capita prescriptions



Other Ongoing Sustainability Applications

Bridging information gaps can reduce food waste

Schema.org collaboration for open standards for reporting Scope 1/2/3 emissions