The Inputs and Outputs that Power America

a MIDS project by Chet Gutwein, Chuqing He, and Max Lee
W209, Fall 2017, Section 2
November 1, 2017

Introduction

- The United States electric grid has rapidly grown: In 2016, 4.08 Billion kiloWatt-hours were produced, 65% from fossil fuels, 20% nuclear, and 15% renewable.

- Nonrenewable energy sources have numerous environmental consequences ranging from carbon and sulfur dioxide to damaging water pollution and deforestation.

- Major milestones like the BP oil spill have raised public awareness and created opportunities for data-utilization as a means to inform and shape people's daily choices.

We Are All Stakeholders

The terms **Sustainability**, **Greenhouse Gases**, and **Clean Coal** have become part of our everyday conversations -- and since Al Gore popularized the phenomenon of climate change a massive environmental movement has been driven.

Now...all Americans must face this issue together. Concerned citizens often ask:

- What are the facts?
- How does the area where I live compare to others?
- Where can my efforts make the most impact?

Data Sources

Emissions & Generation Resource Integrated Database (eGRID): Data provided on power output and emissions for almost all power generated in the United States. Summary data available for 1996-2014, allowing us to introduce a change-over-time factor.
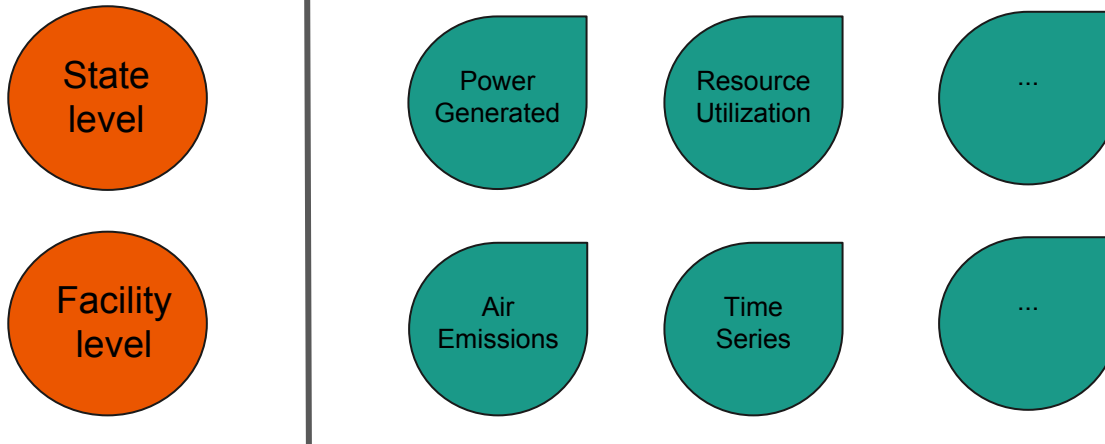

US Energy Information Administration: Data provided on electricity production, sales, across all grids. Majority of data on state level, but some on more granular data at a plant level.

Other data sources that we may implement would look at water usage and waste data.

Tasks

Users will create unique insights with a highly interactive dashboard that will allow dynamic comparisons.

State level

Facility level

Power Generated

Resource Utilization

...

Air Emissions

Time Series

...

From Beginning to Now

Section 2, Chu

- Loading the eGrid dataset
- 18 different Excel files from 1996-2014
- 10 different sheets per Excel file
- Each sheet consist of between 6-30 columns
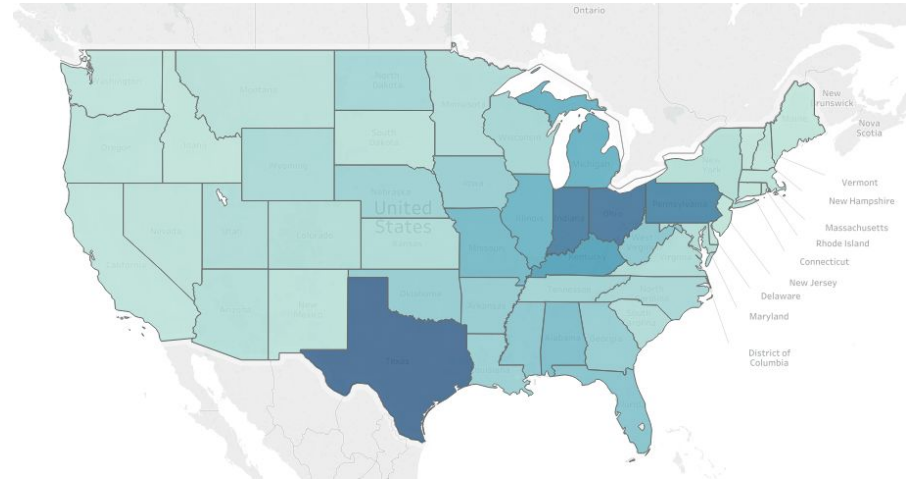
Iteration 1

Section 2, Chu

- Exploratory Data Analysis
- Data Scrubbing to remove/fill-in any empty rows
- Python Pandas to load/aggregate the datasets
- Generate summary statistics on the columns
- Visualization using ggmaps (did not work very well)
- First Visualization using Plotly

Capacity(Megawatts)

US Capacity by State
8000
6000
4000
2000
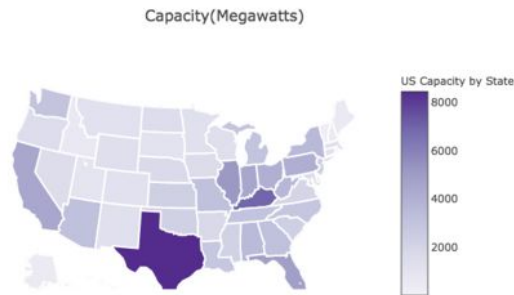
Iteration 2: Tableau

Section 2, Chu

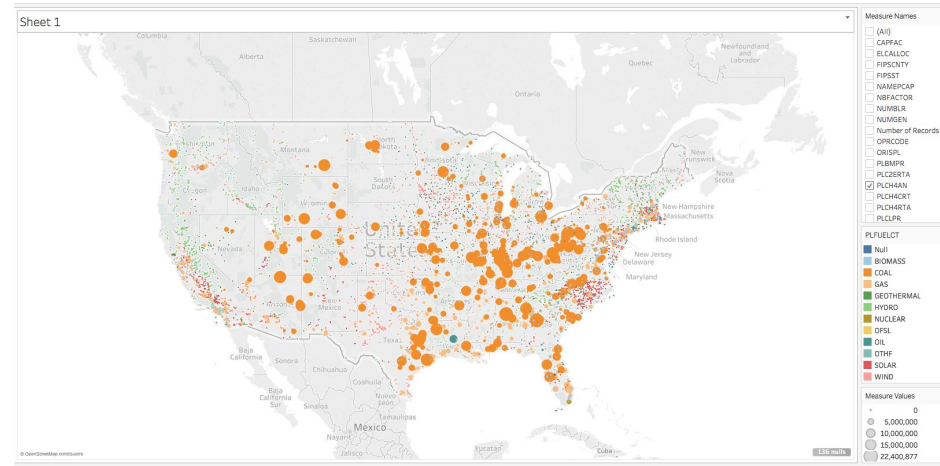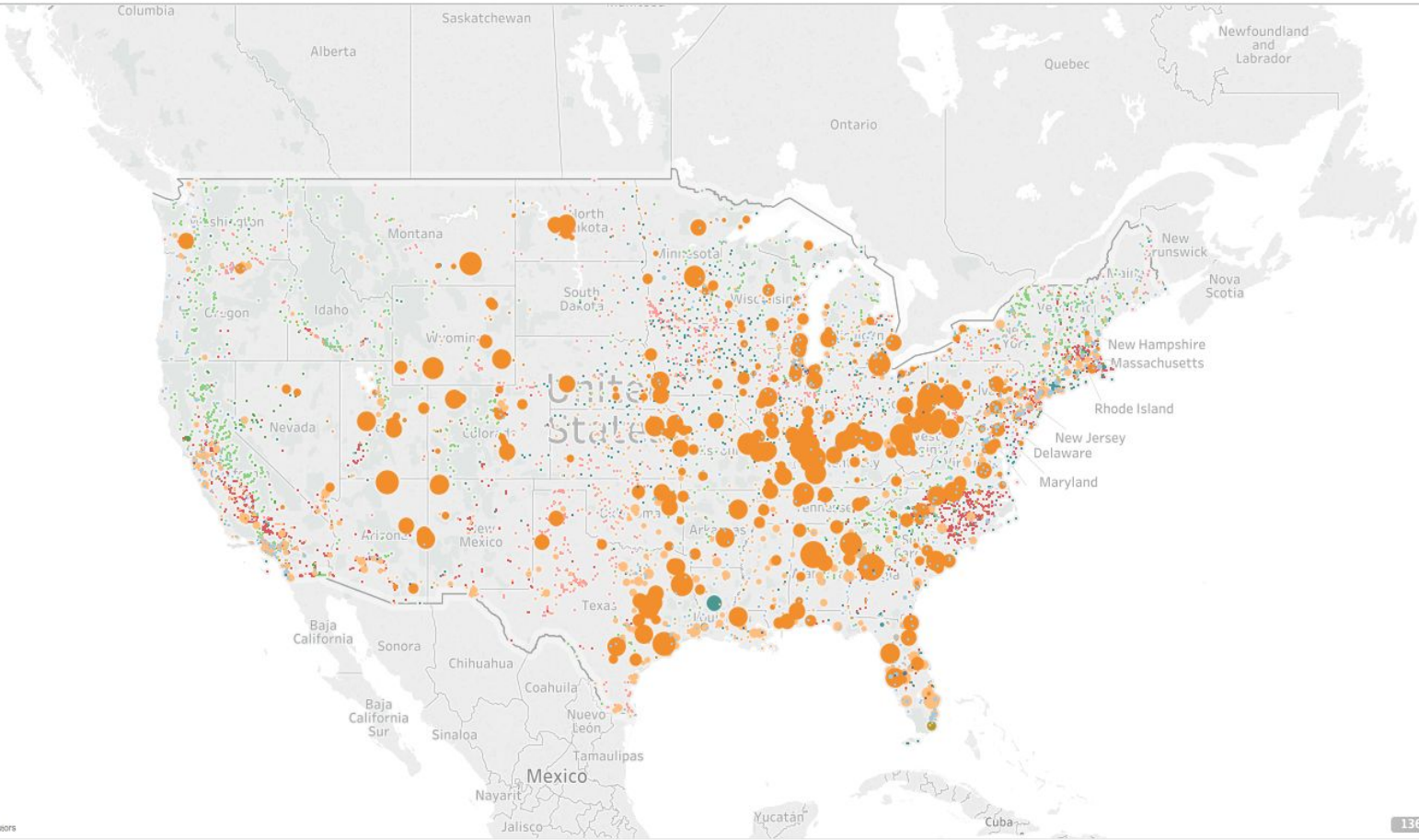-   Visualization using Tableau
-   Aggregate on Watt usage per year

Iteration 3: Improved Tableau

Section 2, Chu

- Improved Visualization using Tableau
- Added different measures
- Added control
- Added Size as a measure instead of heatmap

Sheet 1

Measure Names
- ☐ (All)
- ☐ CAPFAC
- ☐ ELCALLOC
- ☐ FIPSCNTY
- ☐ FIPSST
- ☐ NAMEPCAP
- ☐ NBFACTOR
- ☐ NUMBLR
- ☐ NUMGEN
- ☐ Number of Records
- ☐ OPRCODE
- ☐ ORISPL
- ☐ PLBMPR
- ☐ PLC2ERTA
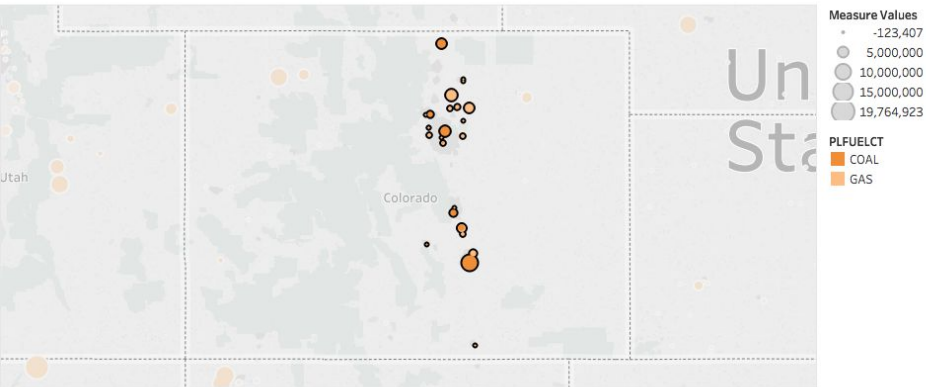- ☑ PLCH4AN
- ☐ PLCH4CRT
- ☐ PLCH4RTA
- ☐ PLCLPR

PLFUELCT
- ■ Null
- ■ BIOMASS
- ■ COAL
- ■ GAS
- ■ GEOTHERMAL
- ■ HYDRO
- ■ NUCLEAR
- ■ OFSL
- ■ OIL
- ■ OTHF
- ■ SOLAR
- ■ WIND

Measure Values
- · 0
- ○ 5,000,000
- ○ 10,000,000
- ○ 15,000,000
- ○ 22,400,877

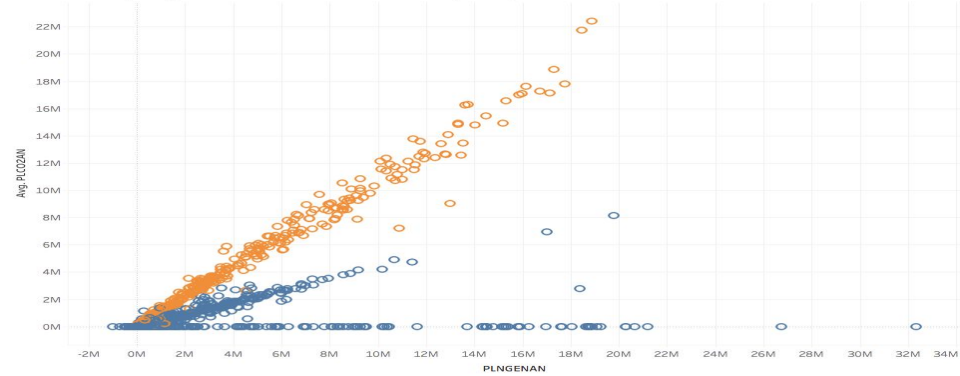136 nulls

© OpenStreetMap contributors

Sheet 1

Map based on LON and LAT. Color shows details about PLFUELCT. Size shows PLNGENAN. Details are shown for PNAME and PLNGENAN. The view is filtered on LON and PLFUELCT. The LON filter keeps non-Null values only. The PLFUELCT filter keeps COAL and GAS.

Measure Values
- -123,407
- 5,000,000
- 10,000,000
- 15,000,000
- 19,764,923

PLFUELCT
- COAL
- GAS

Power Generating Facilities in the U.S.A
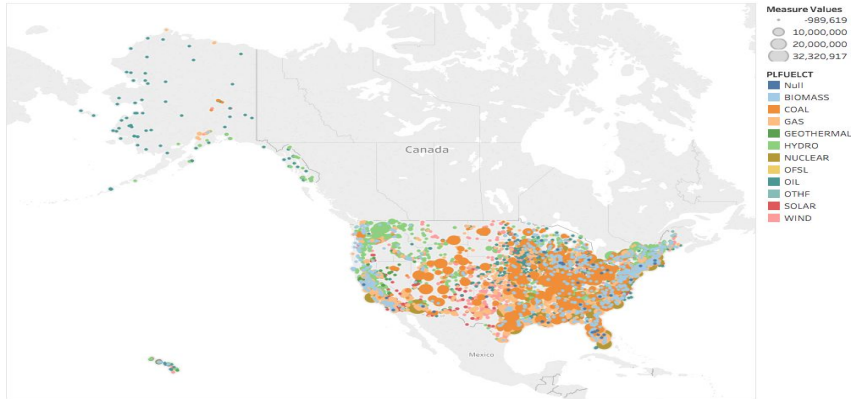
Map based on LON and LAT. Color shows details about PLFUELCT. Size shows PLNGENAN. Details are shown for PNAME and PLNGENAN. The view is filtered on LON and PLFUELCT. The LON filter keeps non-Null values only. The PLFUELCT filter keeps 12 of 12 members.

Measure Values
- -989,619
- 10,000,000
- 20,000,000
- 32,320,917

PLFUELCT
- Null
- BIOMASS
- COAL
- GAS
- GEOTHERMAL
- HYDRO
- NUCLEAR
- OFSL
- OIL
- OTHF
- SOLAR
- WIND

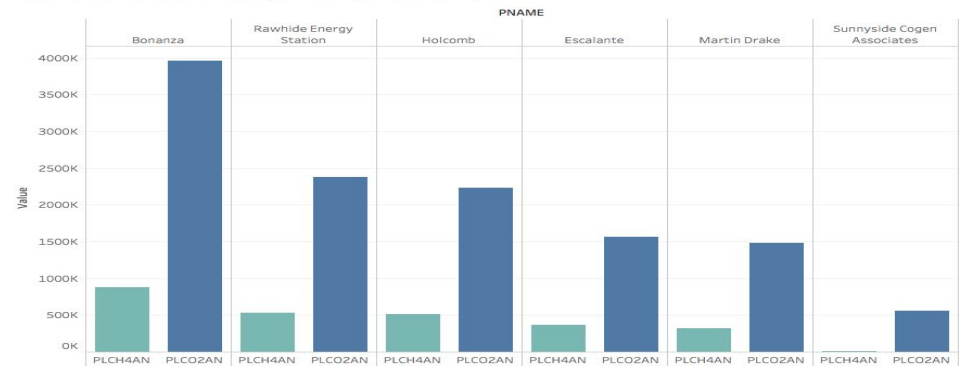### Annual CO2 (tons) vs. Annual Power Generated (MWh)

Sum of PLNGENAN vs. average of PLCO2AN. Color shows details about COALFLAG. Details are shown for PNAME, PLPRMFL and COALFLAG.

COALFLAG
- Null
- Yes

Plant GHG Emissions Comparison (annual tons)

PLCH4AN and PLCO2AN for each PNAME. Color shows details about PLCH4AN and PLCO2AN. The view is filtered on PNAME, which keeps 6 members.

Measure Names
- PLCH4AN
- PLCO2AN

Getting to the Finish Line

- Tableau has provided an easy platform to implement some ideas and test various means of visualization especially in geo-mapping. As we build final webpage, consider the use of d3/JS to allow for highest level of customization.

- Prototyping has focused on small subset of 1 dataset, but goal is to expand temporal dimension (many years) and leverage other fields.

- Aside from basic filtering, how can we better incorporate interactions and transitions?

- Output of other summary statistics, tables, and charts as a means of expressing information aside from maps.

- Add text and develop narrative of what we envision the user to see.

- Mockup: https://balsamiq.cloud/sqtbe/p5izg/r2973
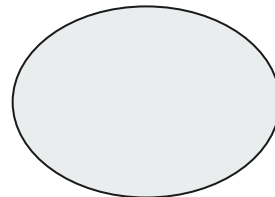
Thank you -- Any Questions??

E-mail: chet_gutwein@berkeley.edu

Slack: @cgutwein

E-mail: chqngh@berkeley.edu

Slack: @chuqing.he

E-mail: maxstlee@gmail.com

Slack: @max_lee