Landfills & Cancer

Group 10: Greg Nations, Zach Liebler, Melissa Kamanzi, Gorkem Varol

Background & Hypothesis

Description:

We'd like to determine whether or not there is a positive correlation between the existence of landfills and the occurrence of cancer in the surrounding populated areas.

Hypothesis:

If there is a concentrated presence of landfills in a given area, then there will be a greater rate of cancer cases in the surrounding areas.

Data Sources & APIs

Data Sources

1. LMOP Database for U.S. Landfills
    a. [https://www.epa.gov/lmop/landfill-technical-data]
2. National Cancer Institute: State Cancer Profiles
    a. [https://www.statecancerprofiles.cancer.gov/incidencerates/]
3. Census.gov Population Data
    a. [https://data.census.gov/cedsci/table]

APIs Used

1. GMaps - Places API
2. GMaps - Geo API

Methods & Tools Used

- Pandas
- Numpy
- Matplotlib
- Jupyter Notebook

Process, Cleanup, & Exploration

- Read in CSV files from each Data Source, separately in our own notebooks.
- Clean the Data, pull out information not relevant to our analysis, average values and create new columns.
- Merge together population and cancer data sets to calculate & identify the total cancer cases per state, rather than cases per 100,000.
- Create new columns to pull in new calculations to help our analysis.
- Plot cancer and population data, as well as the landfill data.
    - Output the new CSVs for use in Gmaps.
- Read in all data sets to visualize the concentration of landfills by location and cancer cases per 100,000 by county.
- Interpret and form our analysis and conclusion.

DataFrame Examples

	state_name	pop_2013	pop_2014	pop_2015	pop_2016	pop_2017	Average_Population
0	AL	4830460	4842481	4853160	4864745	4875120	4853193.2
1	AK	737045	736307	737547	741504	739786	738437.8
2	AZ	6634999	6733840	6833596	6945452	7048876	6839352.6
3	AR	2959549	2967726	2978407	2990410	3002997	2979817.8
4	CA	38280824	38625139	38953142	39209127	39399349	38893516.2

	State	Average_Population	Population_per_100K	Cases_per_100K	Total_Cases
0	AK	738437.8	7.384378	417.2	3080.762502
1	AL	4853193.2	48.531932	451.5	21912.167298
2	AR	2979817.8	29.798178	472.8	14088.578558
3	AZ	6839352.6	68.393526	386.7	26447.776504
4	CA	38893516.2	388.935162	404.6	157363.166545

	State	Average_Population	Population_per_100K
1	AK	738437.8	7.384378
0	AL	4853193.2	48.531932
3	AR	2979817.8	29.798178
2	AZ	6839352.6	68.393526
4	CA	38893516.2	388.935162

	State	Landfill Name	Current Landfill Status	Latitude	Longitude
0	AK	Anchorage Regional Landfill	Open	61.293281	-149.602138
1	AK	Capitol Disposal Landfill	Open	58.352800	-134.494700
2	AK	Central Peninsula Landfill (CPL)	Open	60.447140	-151.103690
3	AK	Palmer Central Landfill	Open	61.590000	-149.210000
4	AK	South Cushman Landfill	Open	64.804760	-147.700850

Problems Overcome

- **Cancer Data**
  - After plotting the cancer cases per 100,000, which is what we had in the raw data set, we realized that it was pretty flat across most states, so we decided to incorporate population data from Census.gov to help calculate the actual number of cases per state. (Shown on slide 10)
  - Our cancer data set ranged from 2013 to 2017, and our population data provided values on an annual basis from 2000-2019. So, we only kept 2013-2017 in scope and took the average as our populations by state.
  - We had issues sorting and merging our two DataFrames (the raw, cleaned DataFrame and the DataFrame that incorporated population data)
- **Gmaps**
  - Which API to integrate?
  - Understanding how smart gmps is? What do you need to tell it and what does it tell you!
  - Visuals tell an emotional story - act on additional senses

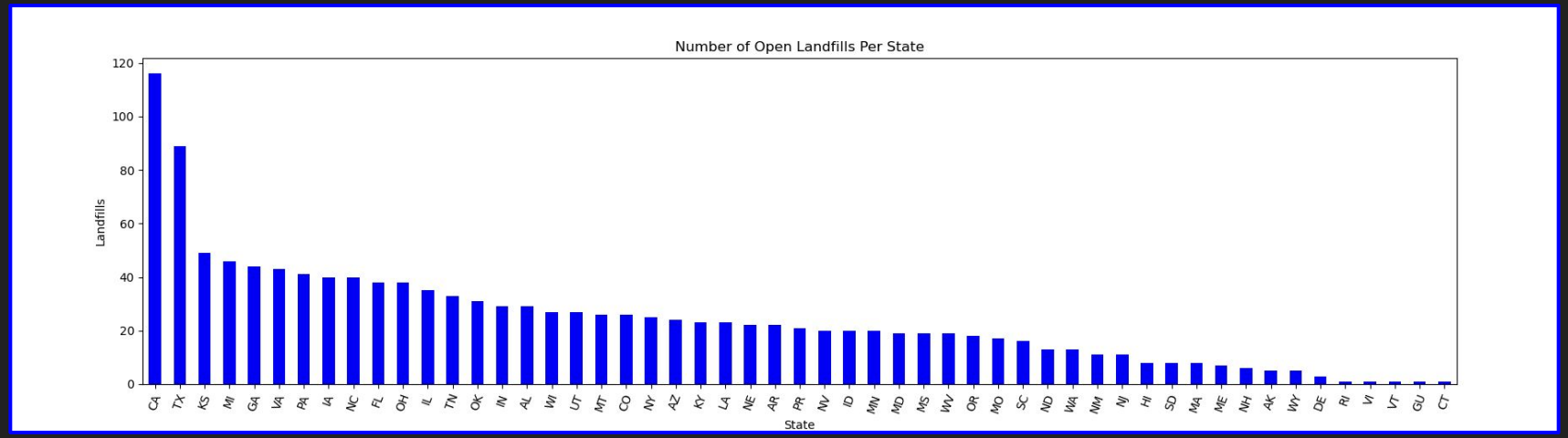
Questions Asked

1. How many landfills are in each state?
2. Which states have the most landfills, and which states have the least?
3. How many cancer cases are there in each state?
4. Visually, what regions in the US have the most open landfills present?
5. Visually, what counties have the most cancer cases per 100,000?
6. Is there a positive correlation between states with a high concentration of landfills, and the number of cancer cases?

How many landfills are there in each state?

Which states have the most, and which states have the least?

Open Landfills Per State

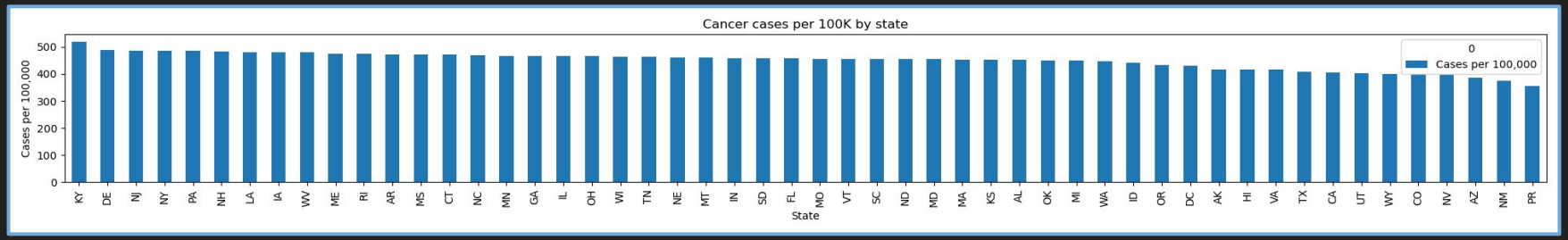


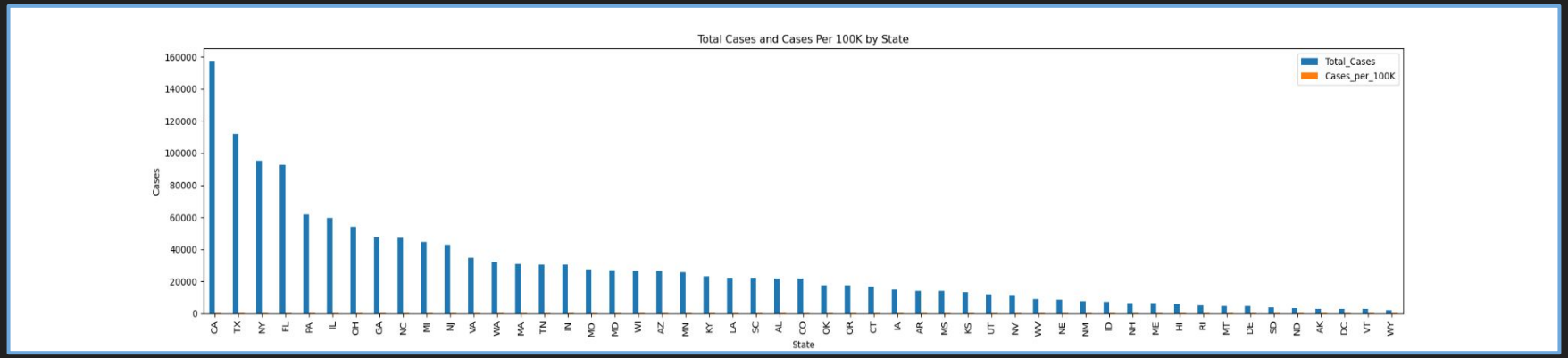
Max - California (116)
Min - Rhode Island, Virginia, Vermont, Connecticut (1)
Median - Puerto Rico (21)

How many cancer cases per 100,000 are there per state?

Layering in population data, how many total cancer cases are there per state?

Cancer Cases per State



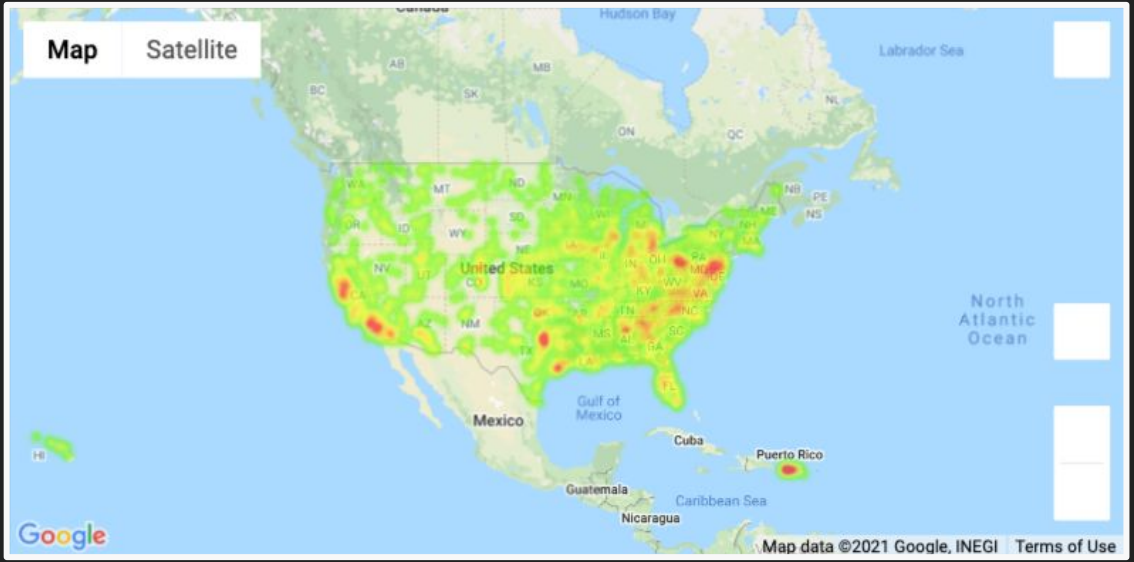


Cancer cases per 100,000 is relatively flat, and did not tell much of a story. We layered in Census.gov population data to help us calculate and identify the number of cancer cases per state.

Visually, what regions in the US have the most open landfills present?

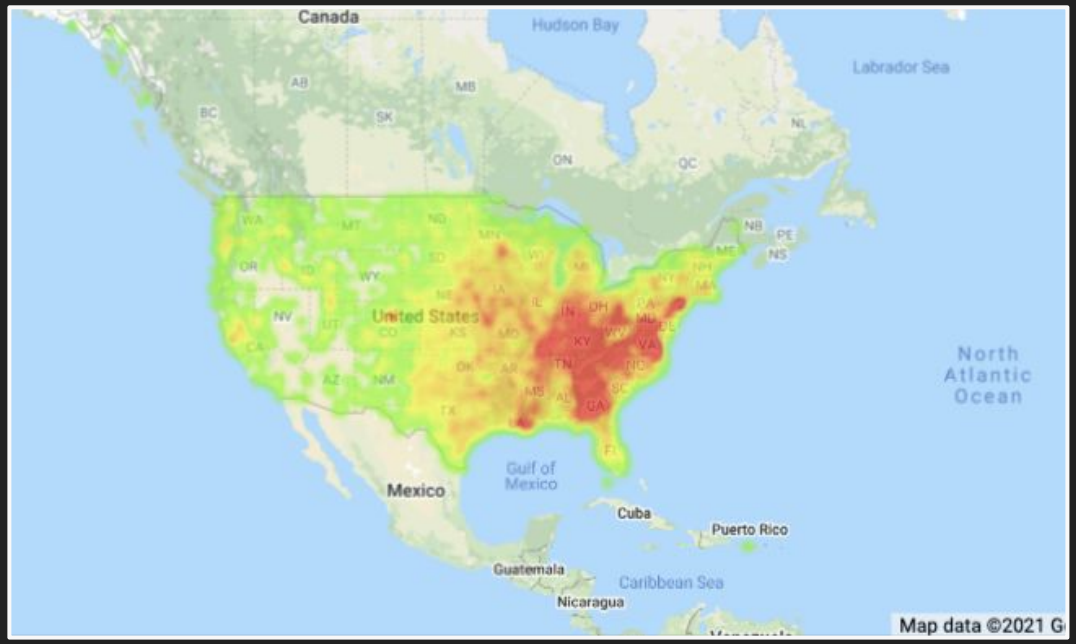
Visually, what counties have the most cancer cases per 100,000?

Open Landfills in the US



The most concentrated areas where landfills are present are in California, Texas, and towards the East Coast along the Appalachian Mountain range.

Cancer Cases Per 100k by County



The most concentrated areas of cancer cases per 100k by County is along the East Coast along the Appalachian Mountain range.

Analysis & Conclusion

Population Bias - Although California and Texas both lead the charge in number of open landfills and total cancer cases, when the cancer data is normalized by cases per 100,000 people, California and Texas are both in the lower half of the data set.

By using heat maps for both open landfills in the US, and cases per 100,000 by county, we're able to see that there are concentrations in both toward the East Coast, near the Appalachian Mountain range.

Analysis & Conclusion

Based on our hypothesis, we could be led to believe that the increased presence of landfills does have influence on the number of cancer cases in the surrounding area, however, once the cancer data is normalized to population data–**we conclude that this hypothesis is false.**

Thank you.

...so Billy should be fine...