

The background features a photograph of a clear blue sky with scattered white clouds, taken from an aerial perspective. A thin white diagonal line runs from the bottom-left corner to the top-right corner, separating the image from the text area.

# Stale Air: Analyzing Historical Air Quality Data in the United States

BANA 295

Javier Orraca | Mark Planas | Yiting Fu

# Research Questions

- Analyze air quality summary statistics by year at the US county level
- How the elements change through time at the county level



# Data Sets

Historical Air Quality Data from  
the Environmental Protection Agency

- Daily and hourly summaries of the elements in the air (Co, O<sub>3</sub>, N<sub>2</sub>O, S<sub>o</sub>2, etc.)
- Other factors (wind, temp, pm10, pm25, etc)
- Date: year, month, day, time



Google  
BigQuery





# Analysis

- Data Selection
- Descriptive Statistics
- Data Visualizations



# Data Selection

- CO
- NO<sub>2</sub>
- O<sub>3</sub>
- Temperature
- Date
- State Name (State Code)
- County Name (County Code)

```
co_1.sort(col('Avg_AQI').desc()).show()
```

state_name	county_name	Avg_AQI	Avg_Value	date
Country Of Mexico California	BAJA CALIFORNIA N...	98.1556886227545	3.555596183183184	1999-11
	Imperial	89.29032258064517	3.1247022419354837	1995-12
Country Of Mexico	BAJA CALIFORNIA N...	82.22489959839358	3.2678619016064268	2000-12
Country Of Mexico	BAJA CALIFORNIA N...	80.85507246376811	3.175846101449275	2002-02
Country Of Mexico	BAJA CALIFORNIA N...	80.1673640167364	3.1400600020920497	1998-01
Country Of Mexico California	CHIHUAHUA STATE	77.33333333333333	2.7795186916666665	1998-11
	Imperial	76.9	2.721872316666667	1995-11
Country Of Mexico	BAJA CALIFORNIA N...	76.76129032258065	3.054804090322581	1999-12
Country Of Mexico	BAJA CALIFORNIA N...	76.0460251046025	2.899022043933054	1998-11
Country Of Mexico Oregon	BAJA CALIFORNIA N...	75.64556962025317	2.874188529535865	1997-11
	Josephine	71.74193548387096	4.064534903225807	1991-01
Country Of Mexico	BAJA CALIFORNIA N...	71.26728110599079	2.7685917304147467	1999-01
Country Of Mexico Oregon	CHIHUAHUA STATE	71.05	2.408870741666666	1997-11
	Jackson	69.62295081967213	3.9828837049180317	1991-01
Washington		Yakima	68.80645161290323	3.9224633548387096
Country Of Mexico Montana	BAJA CALIFORNIA N...	67.54435483870968	2.6546290564516126	1998-12
	Missoula	66.12903225806451	3.879083129032258	1992-01
Country Of Mexico Nevada	BAJA CALIFORNIA N...	65.98373983739837	2.5543170365853665	1997-12
	Clark	65.96774193548387	3.0967487258064517	1990-12
Washington		Snohomish	65.33333333333333	3.7086163500000002
				1990-09

# Descriptive Statistics

```
co = co.withColumn("date",
date_format('date_local',
"yyyy-MM"))

co_1 = 

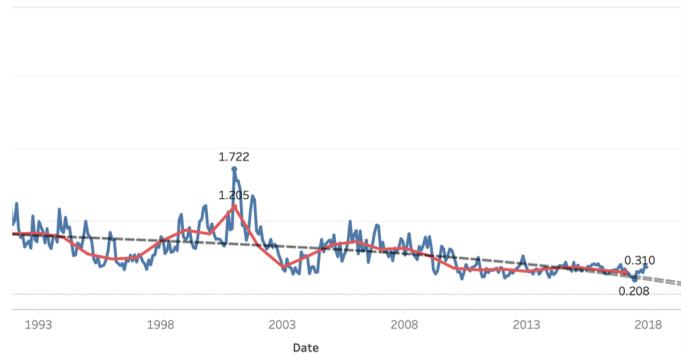
co.groupBy('state_name',
'county_name', 'date')

.avg('aqi', 'arithmetic_mean')

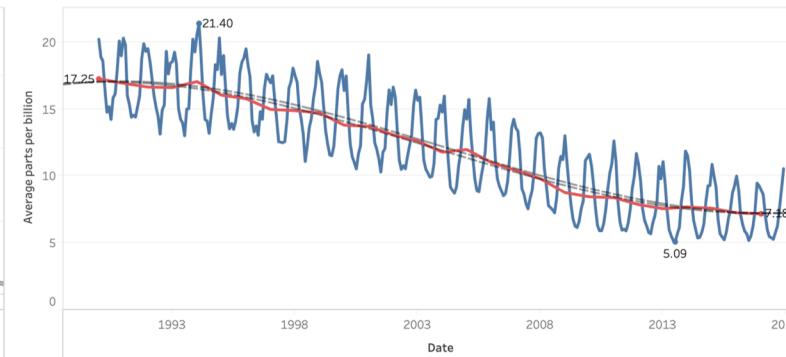
.select('state_name',
'county_name',

col("avg(aqi)").alias("Avg_AQI"),
col("avg(arithmetic_mean)").alias("Avg_Value"), 'date')
```

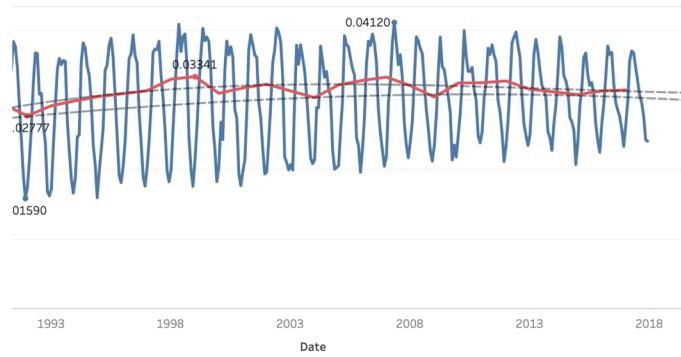
Carbon Dioxide Levels - All



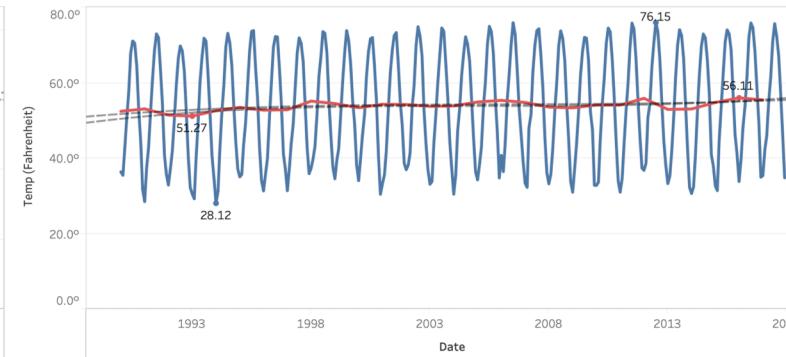
Average Nitrogen Dioxide Levels - All



Ozone Levels - All



Average Temperature Levels - All



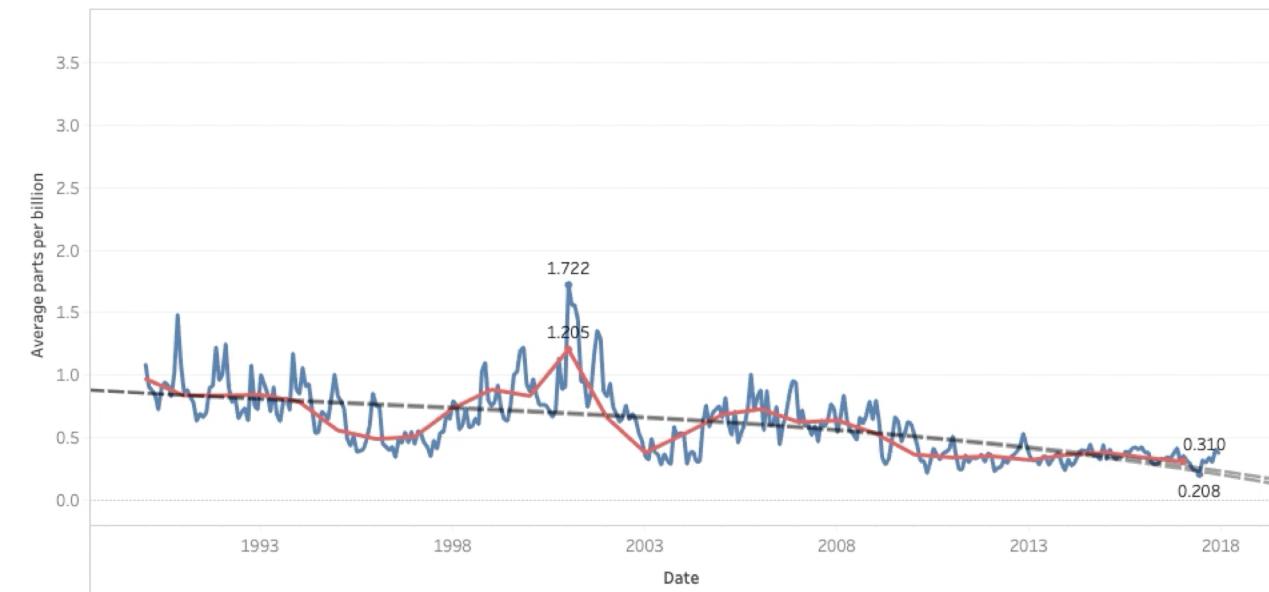
# Data Visualization

The level changes of

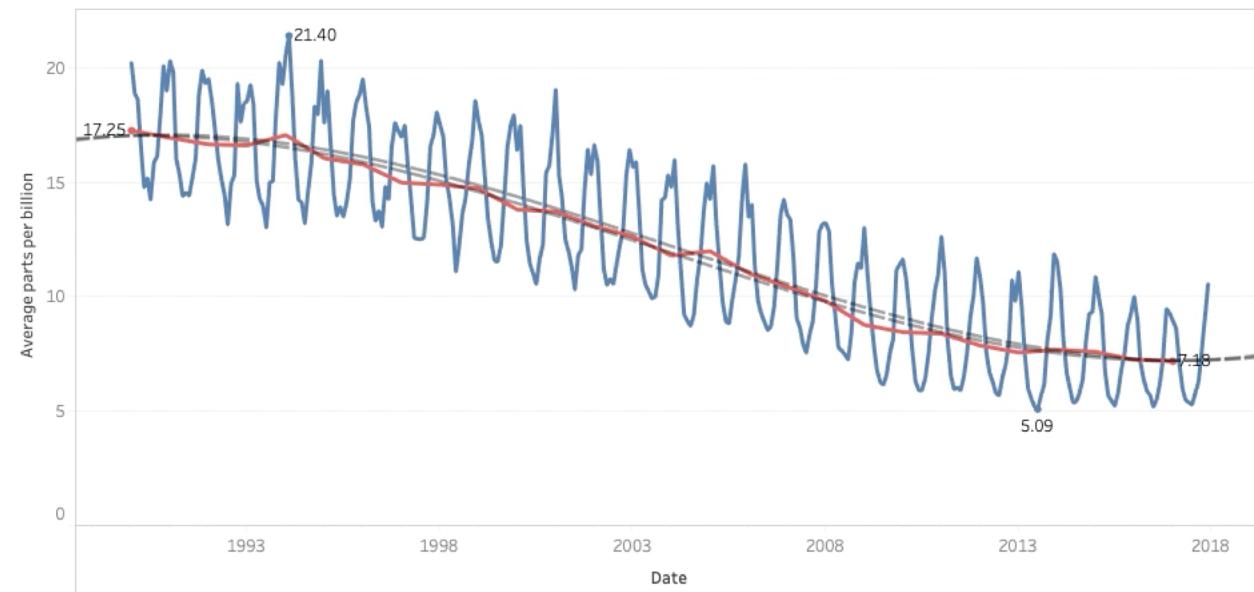
- Carbon Dioxide
- Nitrogen Dioxide
- Ozone
- Temperature

throughout time...

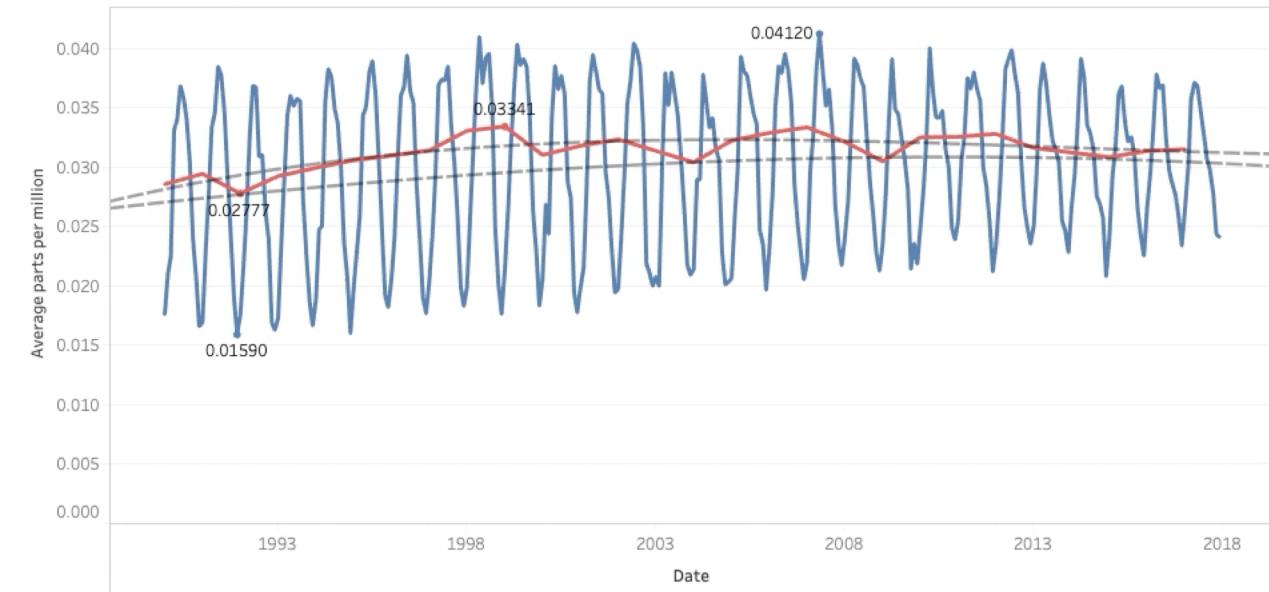
Average Carbon Dioxide Levels - All



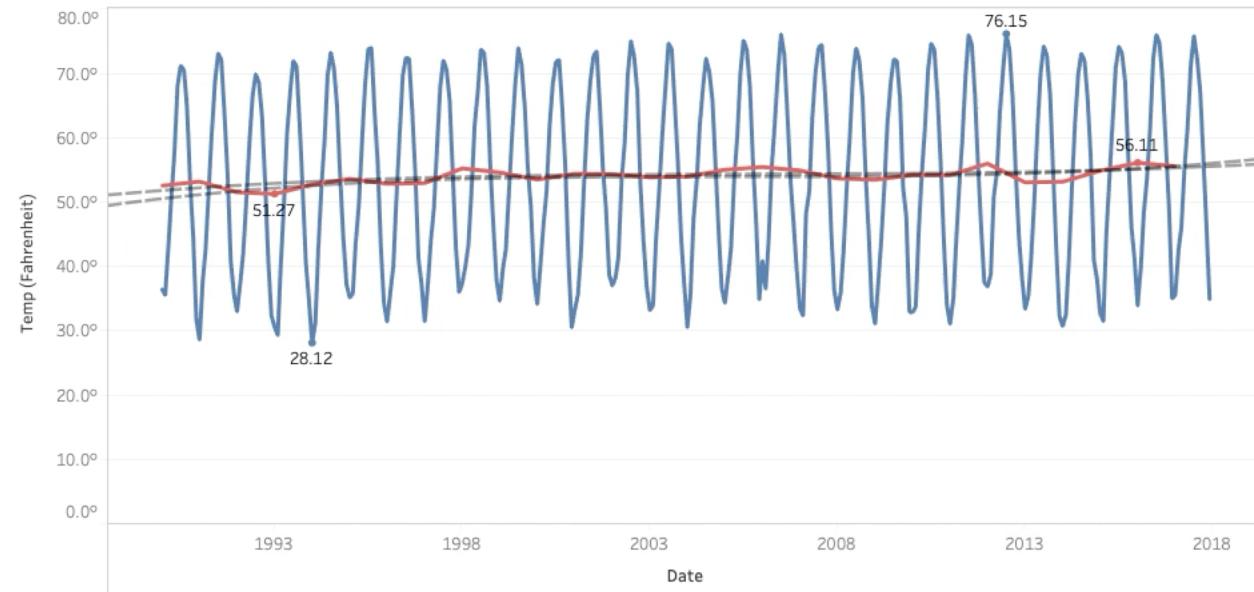
Average Nitrogen Dioxide Levels - All



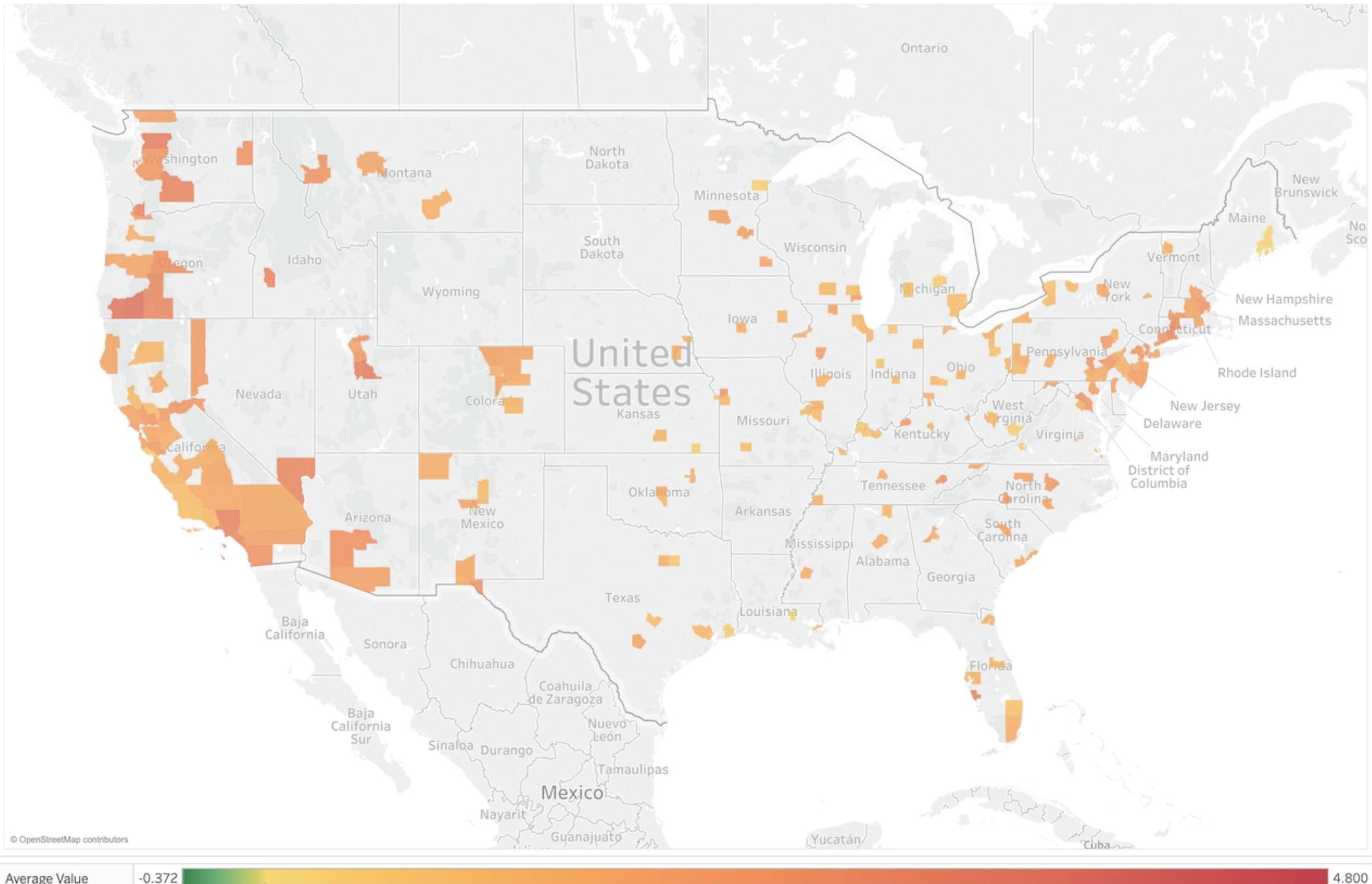
Average Ozone Levels - All



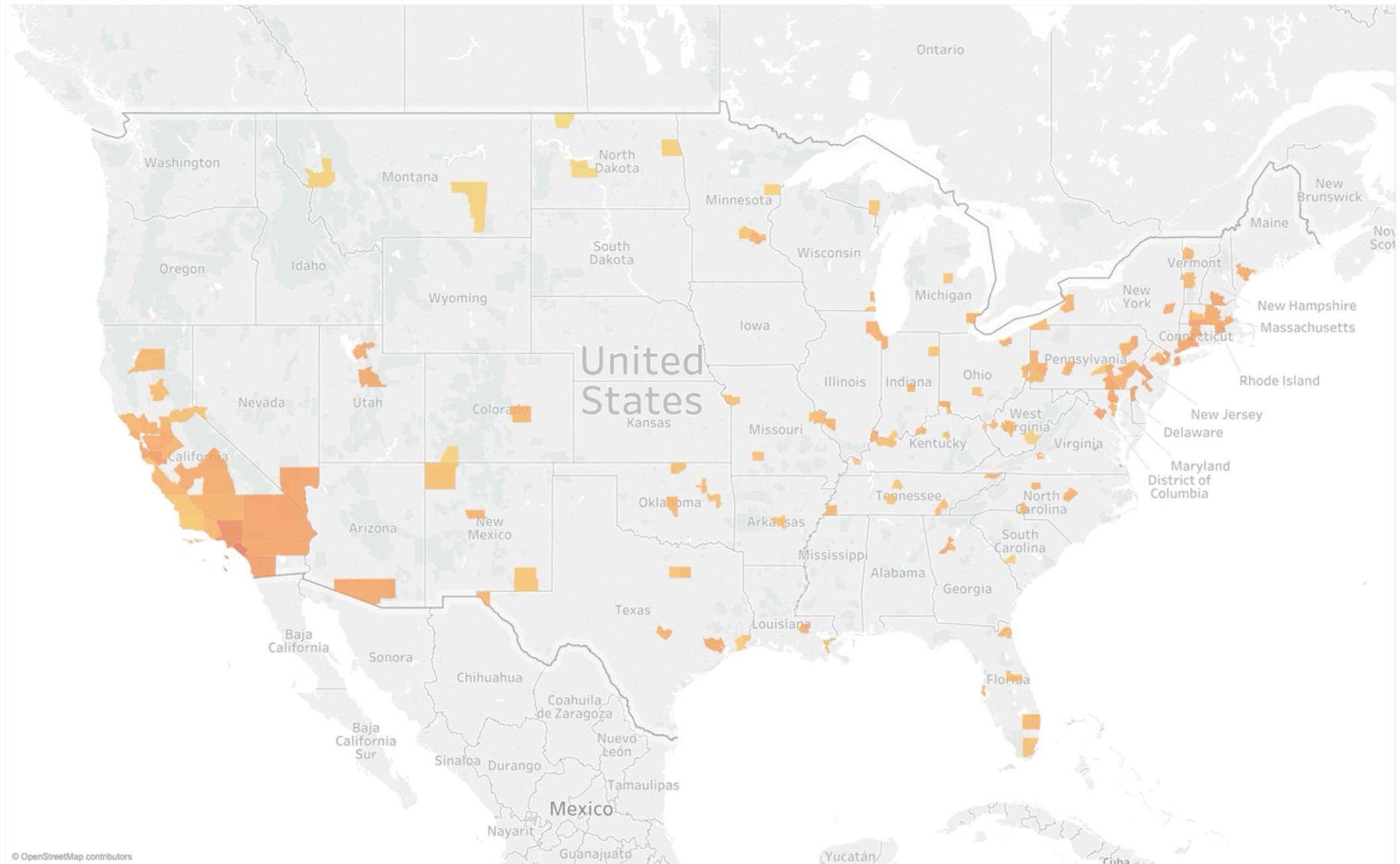
Average Temperature Levels - All



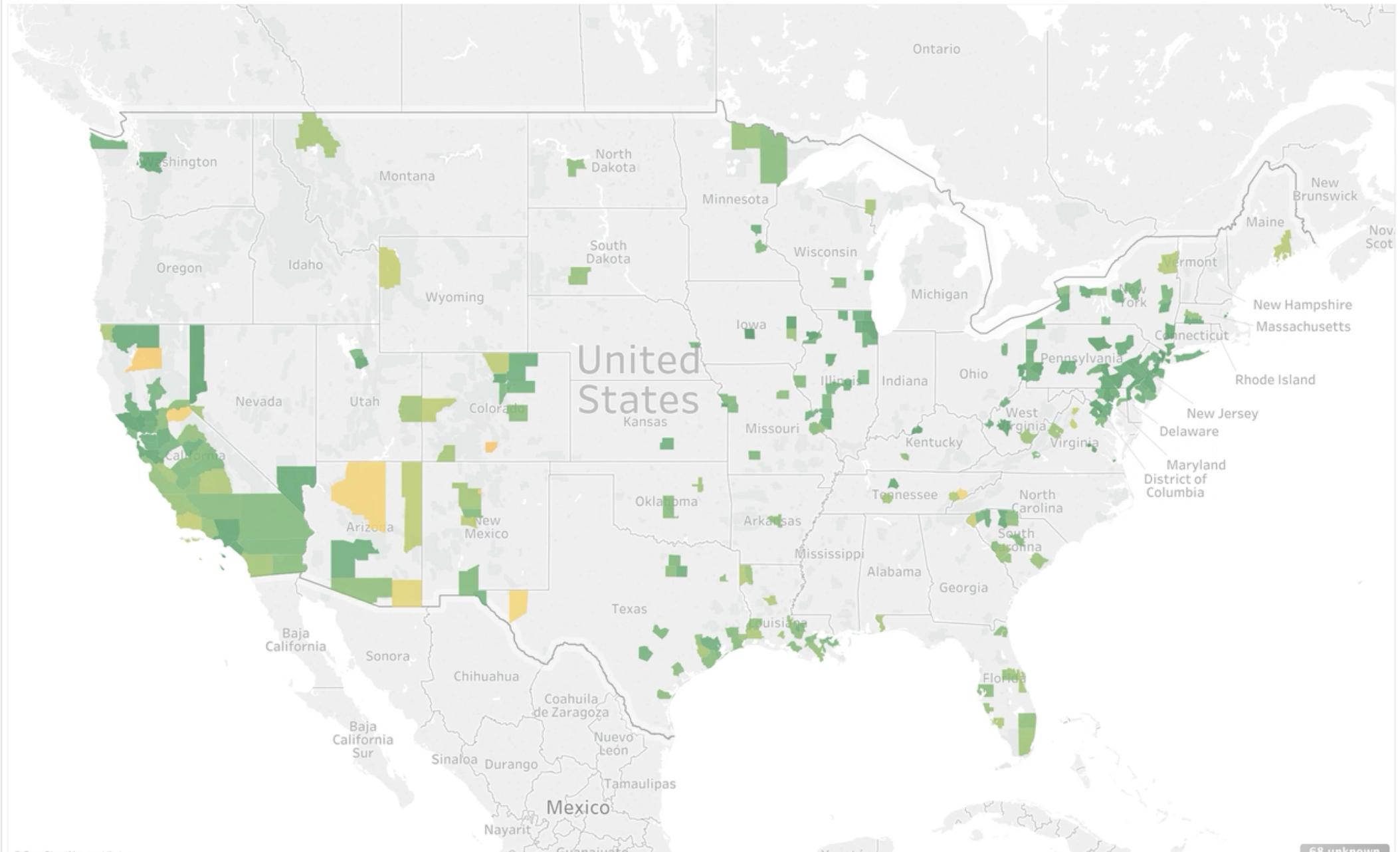
## Carbon Dioxide Levels: January 1990



## Nitrogen Dioxide Levels: January 1990



## Ozone Levels: January 1990

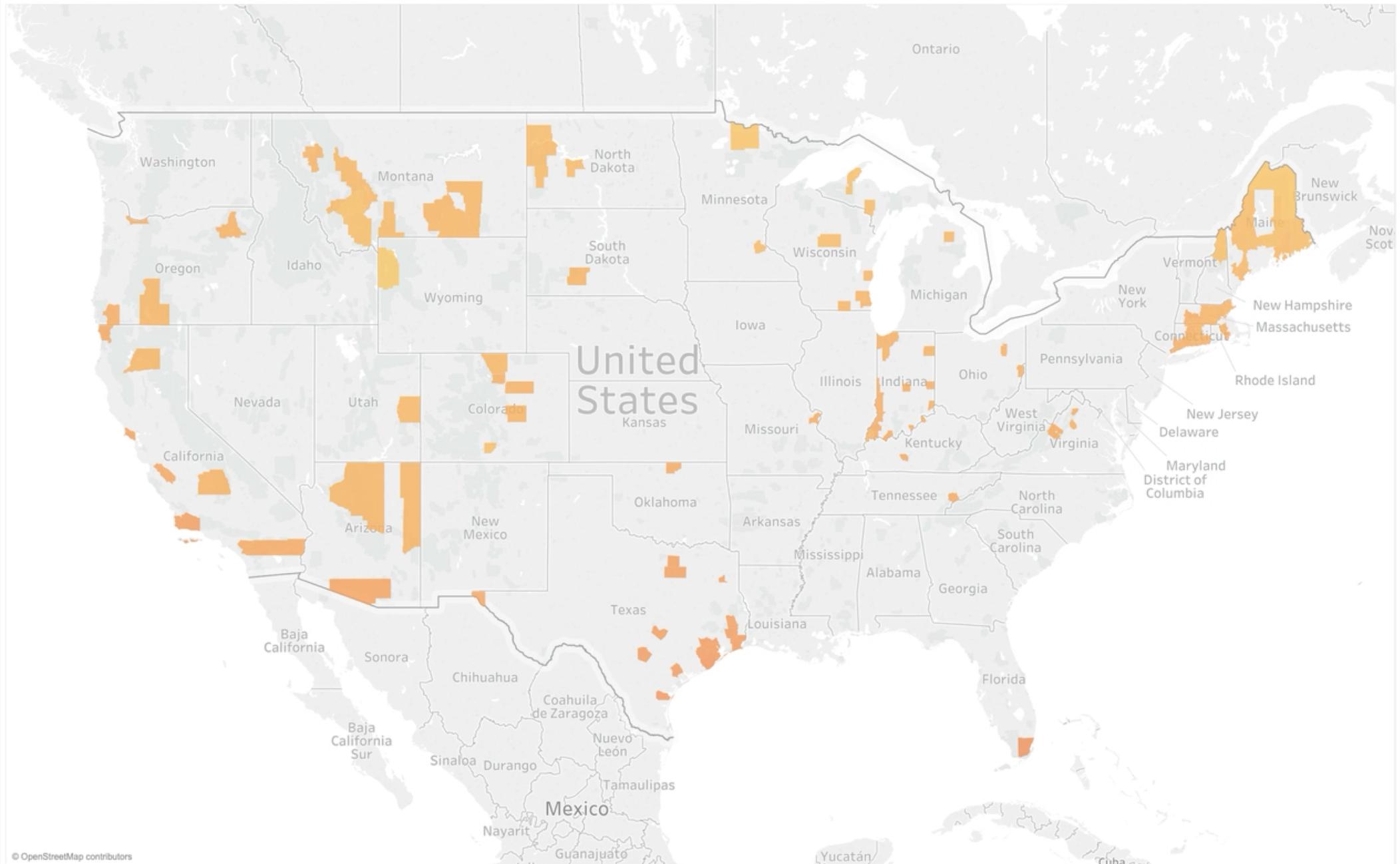


Average Value

0.00000

0.07603

## Temperature Levels: January 1990





# Conclusion

- While carbon dioxide and nitrogen dioxide levels are decreasing, temperature has remained roughly stable.
- Ozone has increased since 1990, and while this is good for the stratosphere, ozone in the air we breathe is harmful to our health.
- Breathing ozone can trigger a variety of health problems including chest pain, coughing, throat irritation, and airway inflammation.

# Challenges

- Finding the most meaningful data
- Kaggle kernels do not support Spark
- Google Cloud Platform could improve accessibility of BigQuery data through Google Colab
- Spark and PySpark installation is worse than the syntax alone



# Key Process Findings

- **PySpark** allowed for computational efficiencies not typically seen when manipulating and modeling Pandas data frames
- To collaborate and code at the same time, we utilized several **Google Cloud Platform** tools to gain access to **BigQuery** data, and we relied on a **Google Colab** for scripting, a cloud-based IDE where we could initiate Spark clusters
- Our team was very surprised to discover that we could run our entire **Google Colab / PySpark** code, including the creation of **Plotly** visualizations, without hitting the 12GB RAM GPU limitation imposed on free, public Google Colab accounts



Thank you

# THANK YOU!

