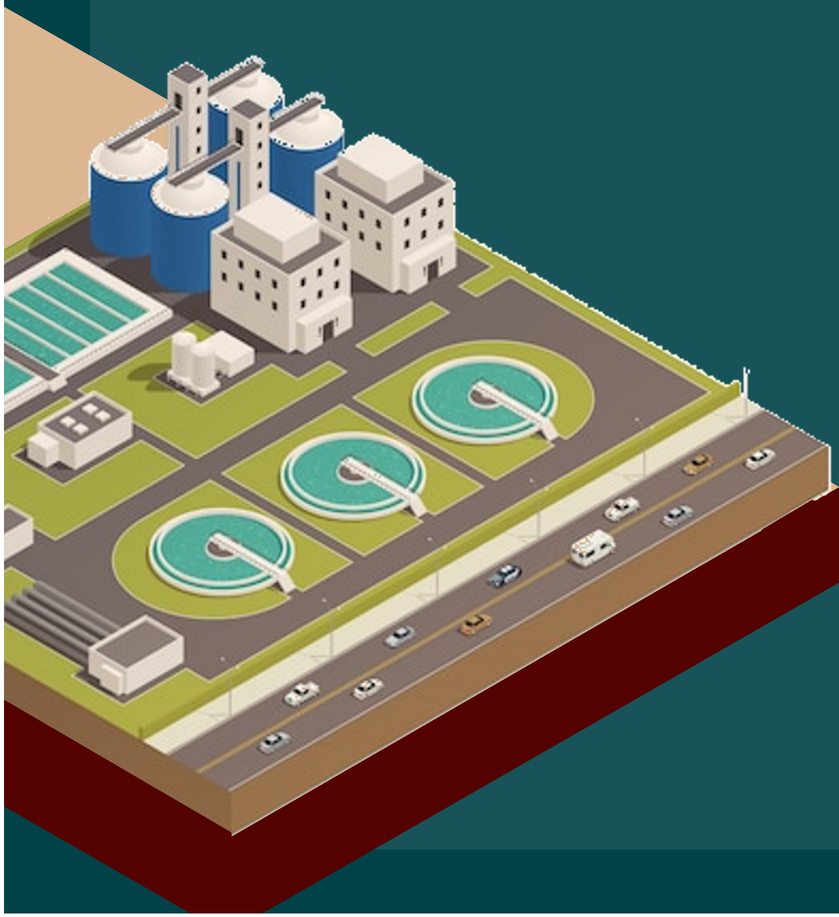


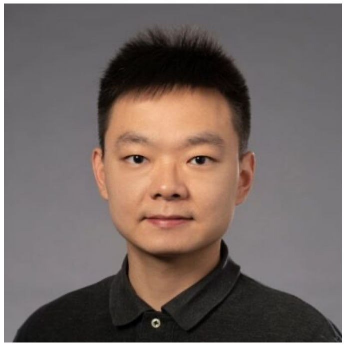
Analyzing the Scope and Scale of Solar Energy Adoption at Wastewater Treatment Plants in the US

Capstone Symposium
March 2024

Our Team

Pooja Kabber

Sukhpreet Sahota



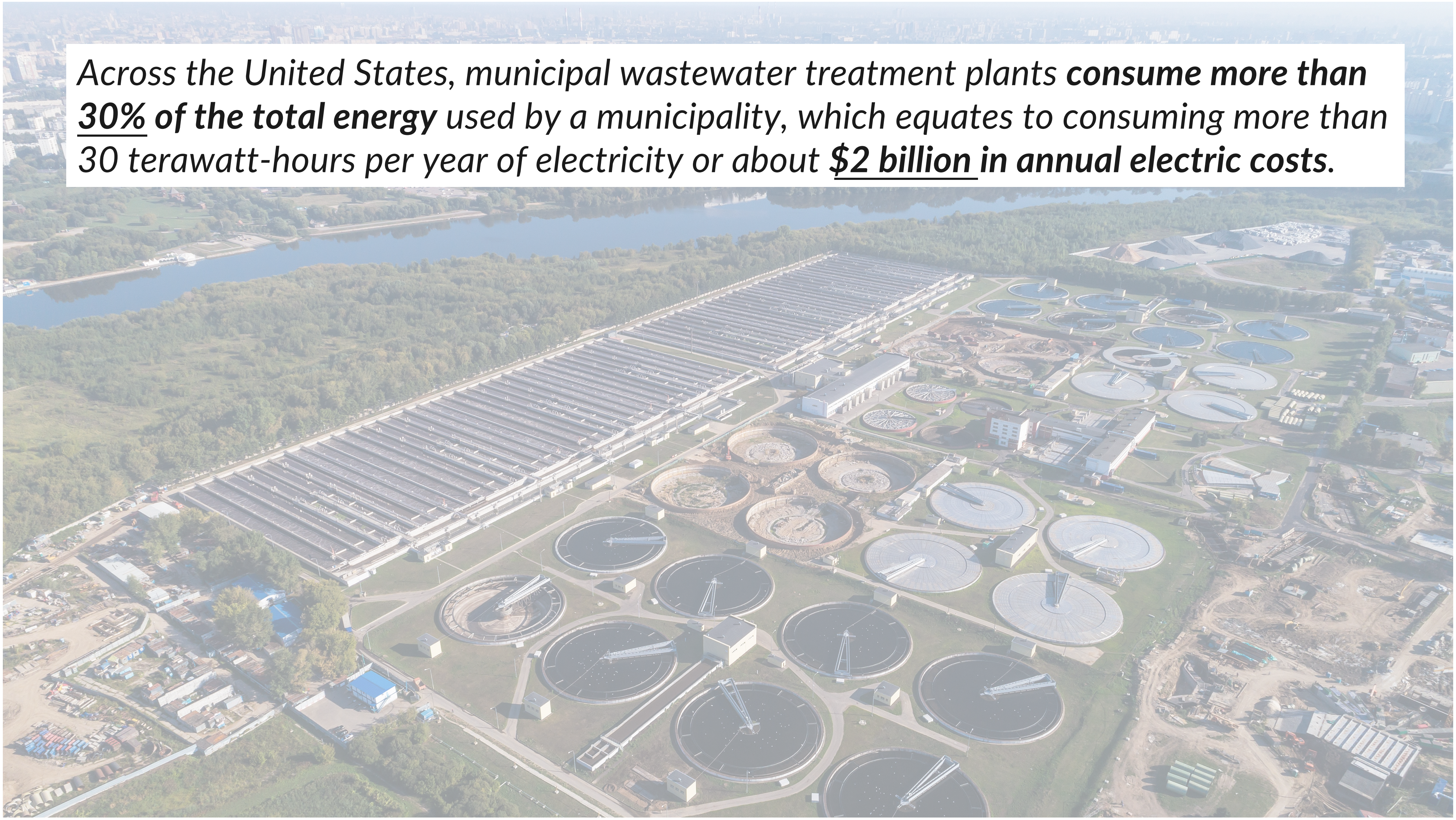
Dingkun Yang

Yuanjing Zhu

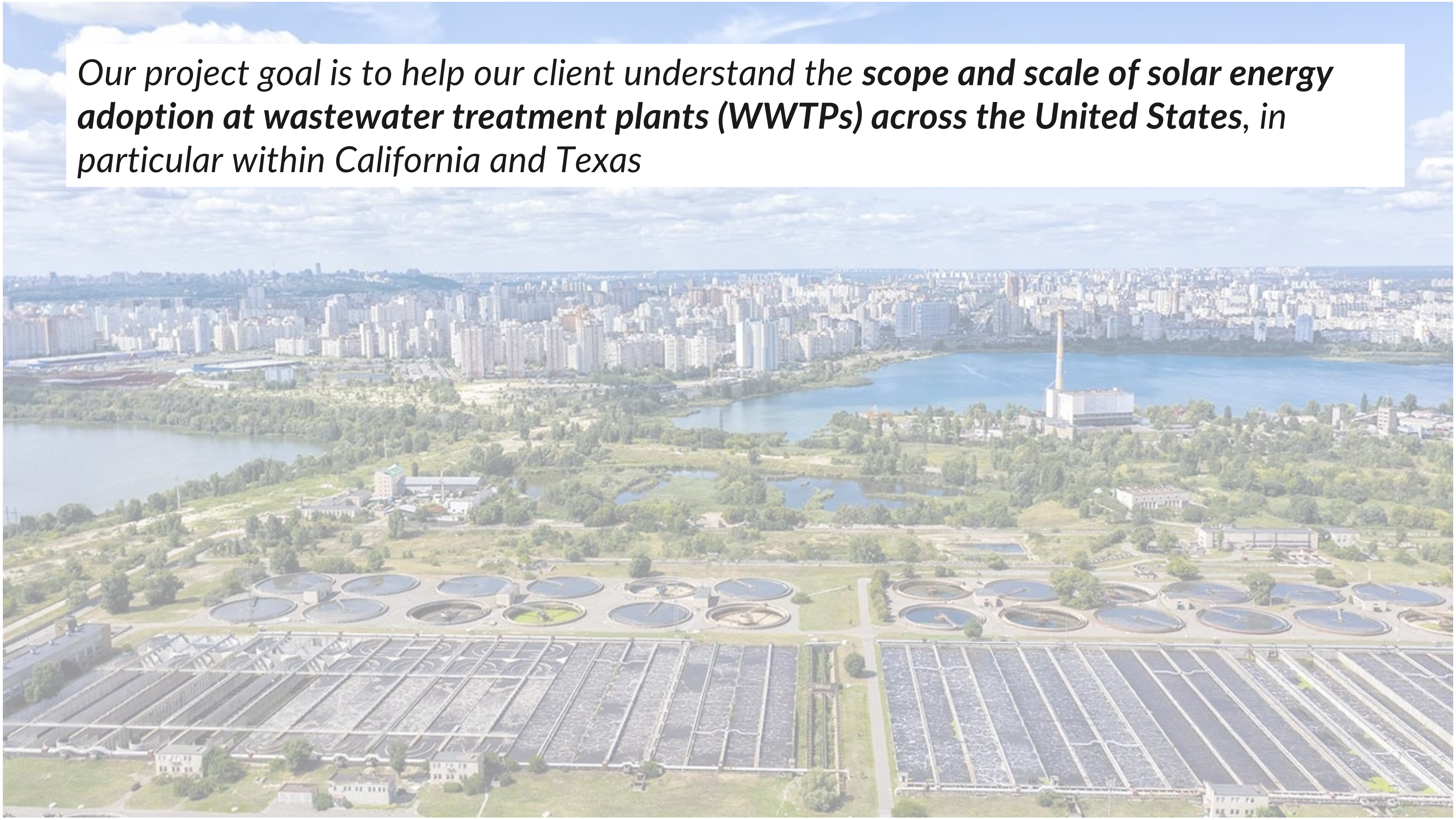
Our Capstone Partners

Duke
NICHOLAS INSTITUTE
for ENERGY, ENVIRONMENT
& SUSTAINABILITY

PennState
Institute of Energy
and the Environment

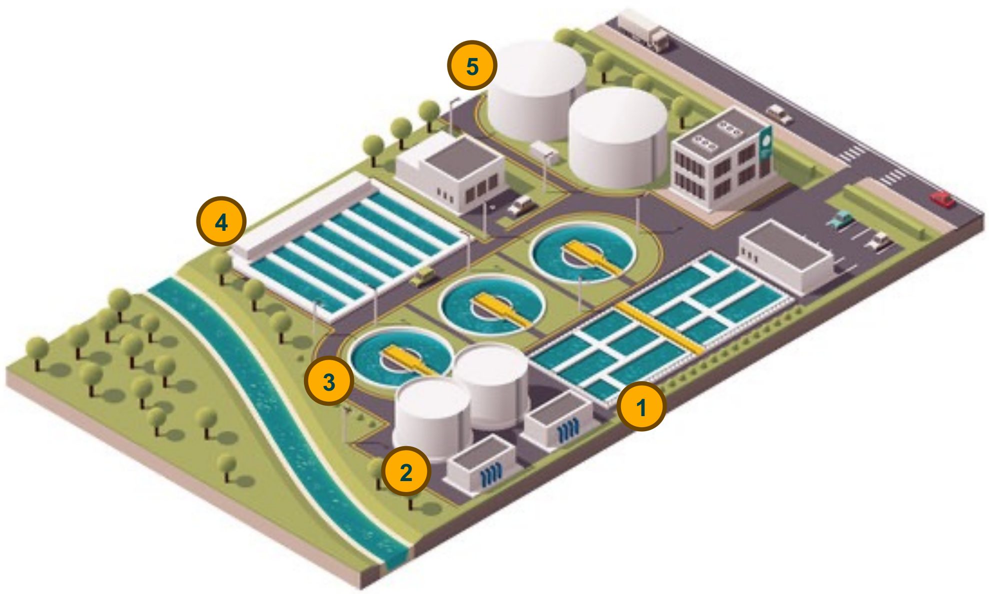


*Across the United States, municipal wastewater treatment plants **consume more than 30% of the total energy** used by a municipality, which equates to consuming more than 30 terawatt-hours per year of electricity or about **$2 billion in annual electric costs**.*



*Our project goal is to help our client understand the **scope and scale of solar energy adoption at wastewater treatment plants (WWTPs) across the United States**, in particular within California and Texas*

What is a Wastewater Treatment Plant (WWTP)?



Wastewater treatment plants are facilities designated to remove contaminants from sewage from their respective surrounding communities and environments

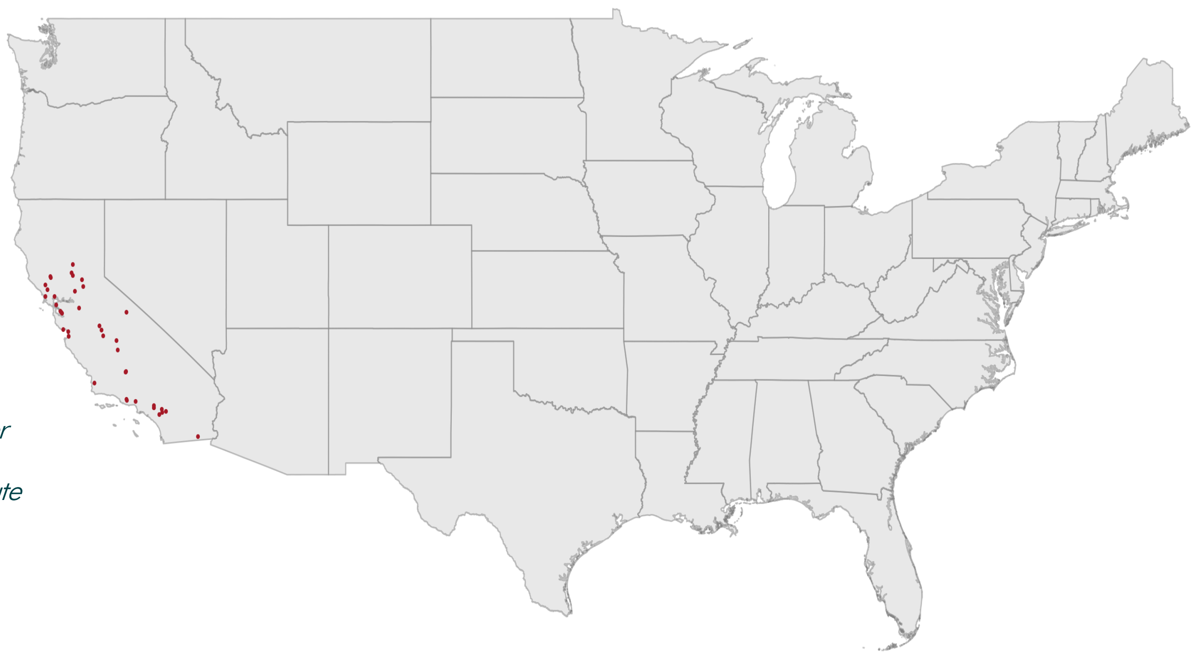
Key Parts of a Wastewater Treatment Plant:

1. Aeration Basins
2. Granular Reactors
3. Clarifiers
4. Filter lagoons
5. Digesters

Identifying Wastewater Treatment Plants with Solar

Data Source and Analysis

Data Sources



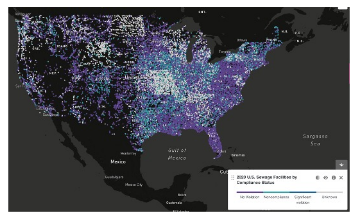
Approved List of 40 Solar Verified WWTP

An initial verified list of 40 wastewater treatment plants that utilized solar energy provided by Penn State Institute of Energy and the Environment

Data Sources

OpenStreetMap

A free, open, and crowdsourced geographic database updated and maintained by a community of volunteers via open collaboration.



Environmental Protection Agency (EPA)

An government published data set containing information relating to location and facility identification from the EPA's Facility Registry Service (FRS) from 2020

HydroSheds HydroWaste

An open-source spatially explicit global database of 58,502 wastewater treatment plants (WWTPs)

List of Aggregated WWTPs Include Sites That Aren't Actual WWTPs

OpenStreetMap

EBMUD Wastewater Treatment Plant, CA

City of Mt. Enterprise Wastewater Treatment Plant, TX

List of Aggregated WWTPs Include Sites That Aren't Actual WWTPs

<u>Environmental Protection Agency (EPA)</u>

Haskell St WWTP, TX

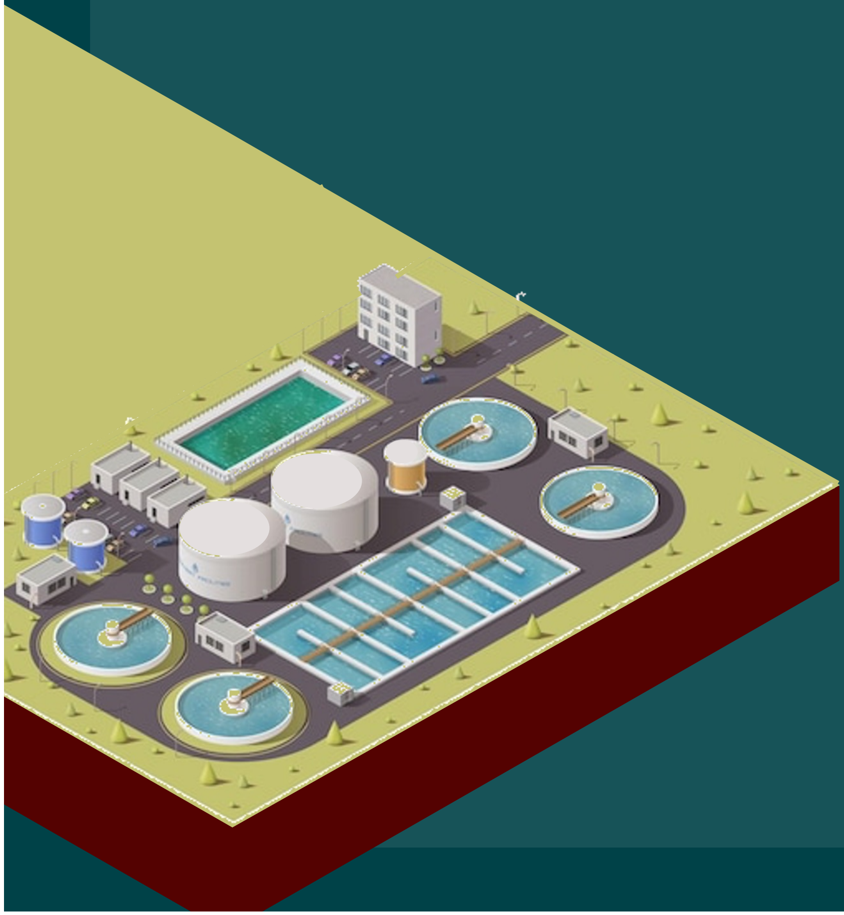
Tahoe Truckee WWTP, CA

List of Aggregated WWTPs Include Sites That Aren't Actual WWTPs

<u>HydroSheds HydroWaste</u>

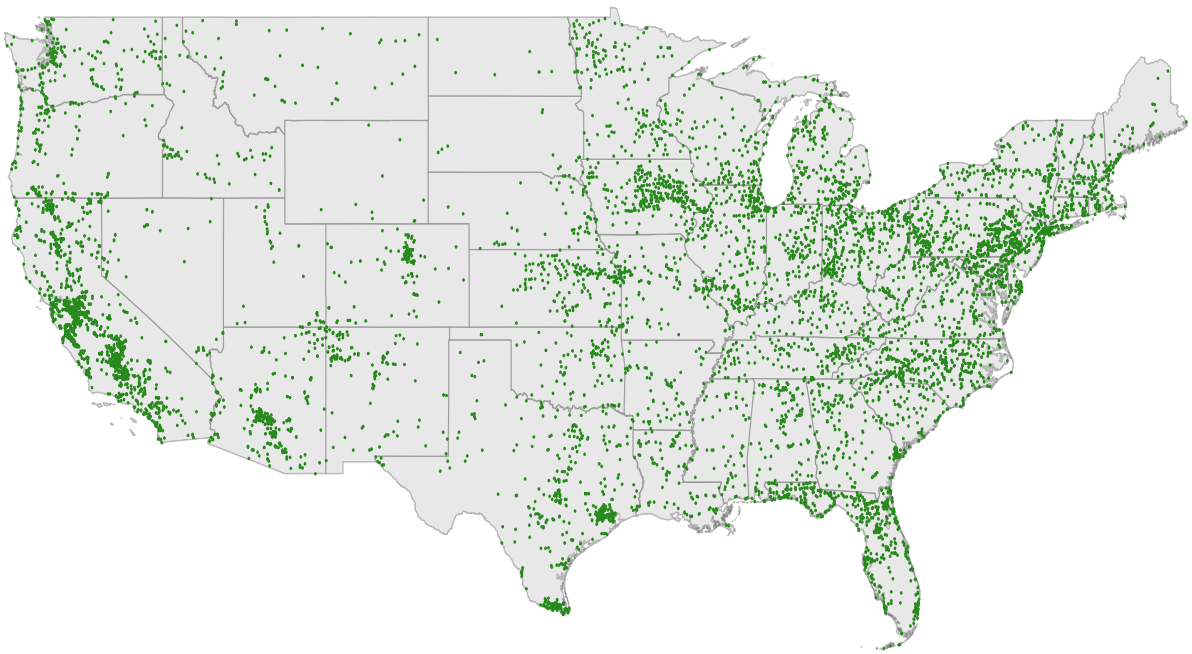
Woodland WWTF, CA

Hart WWTP, TX



How large is this problem?

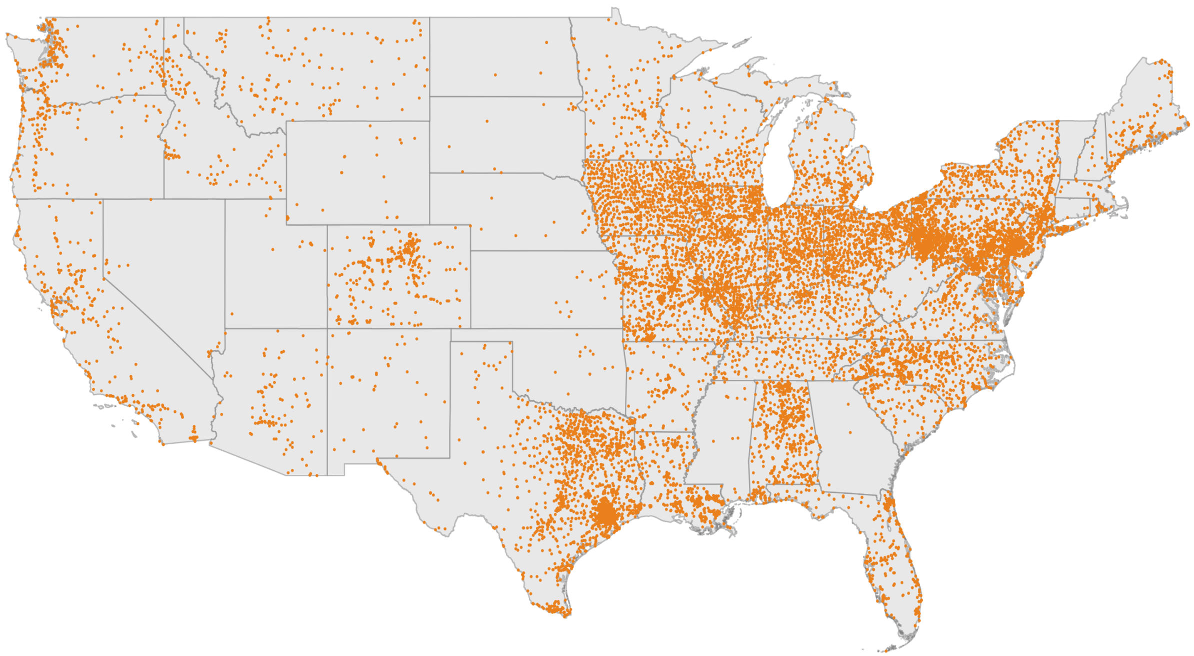
Data Source Analysis



OpenStreet Map

14,282 WWTPs

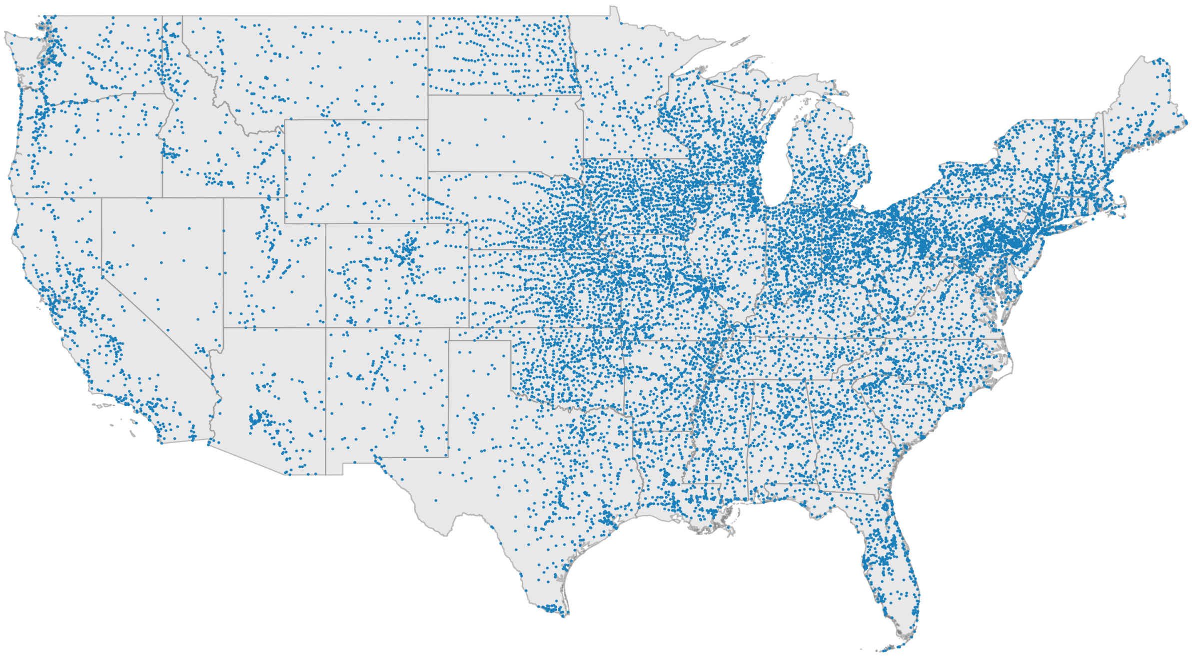
Data Source Analysis



Environmental Protection Agency (EPA)

14,327 WWTPs

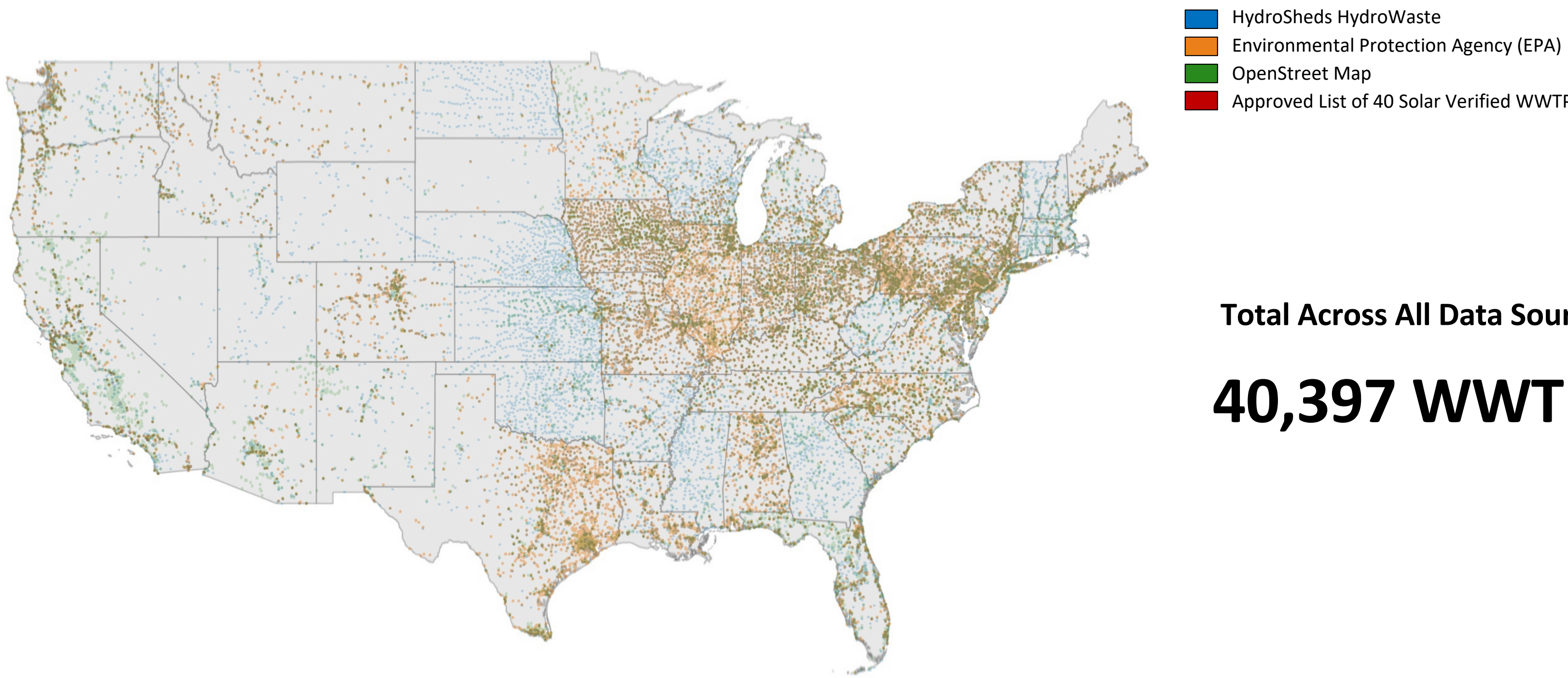
Data Source Analysis



HydroSheds HydroWaste

14,748 WWTPs

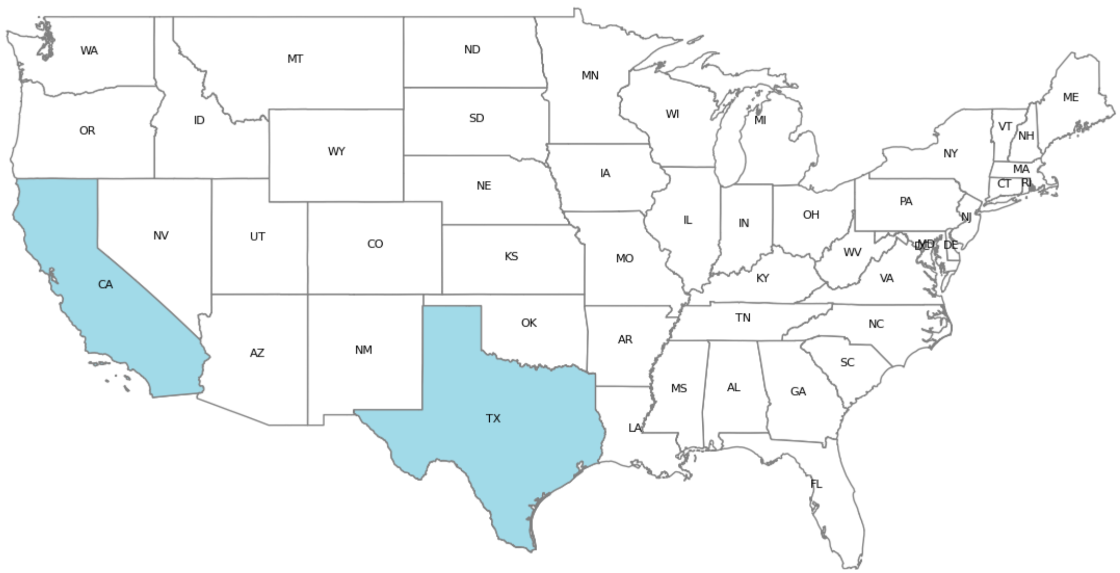
Data Source Analysis



Total Across All Data Sources

40,397 WWTPs

To Tackle the Challenges

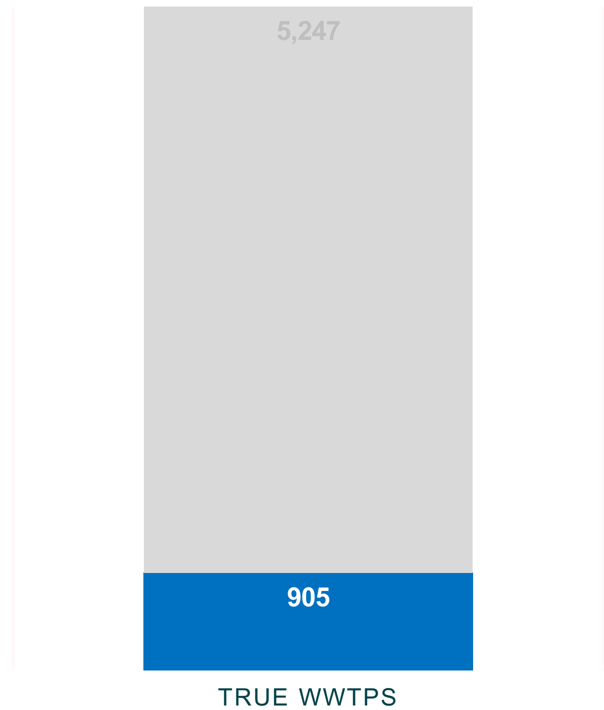


Using California and Texas as our initial pilot states:

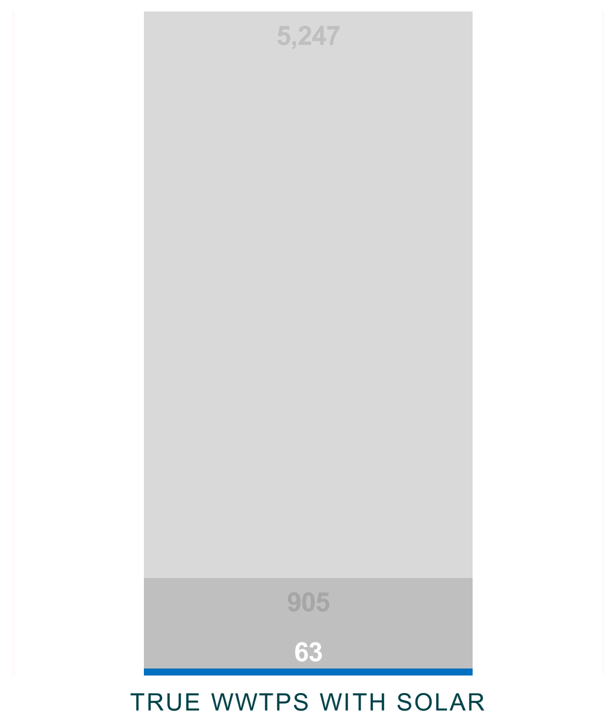
1. Client is interested in those respective states
2. Large population states would contain more WWTPs
3. Demographic and economic diversity

For California and Texas, 5247 image candidates to verify across all data sources

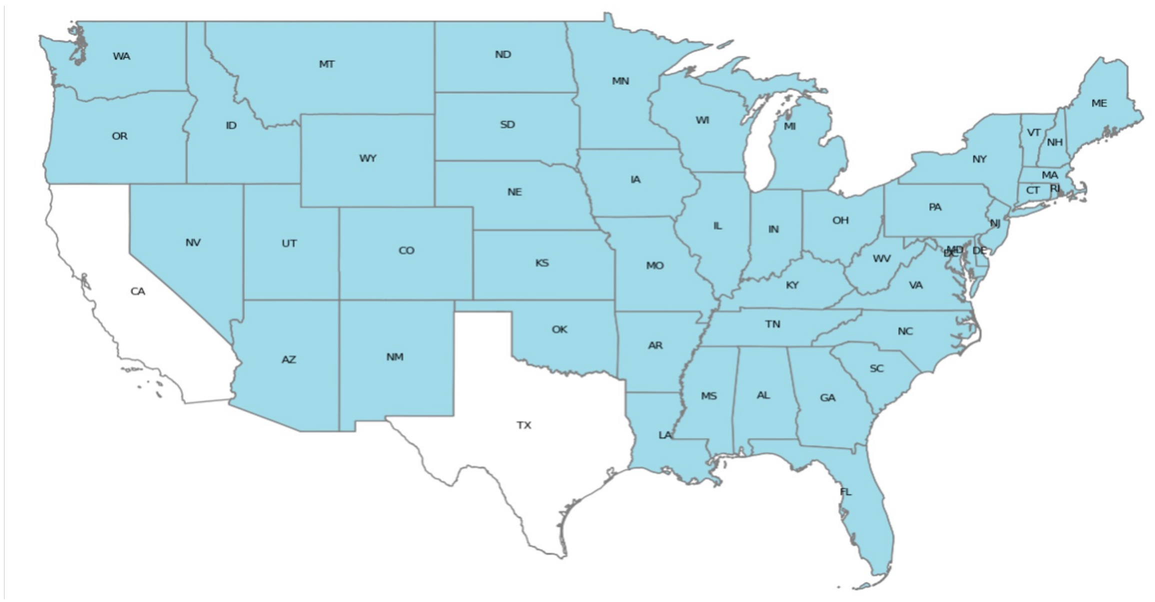
17.25% of image candidates have Wastewater Treatment Plant in them for California and Texas



6.96% of Wastewater Treatment Plants have solar adoption in California and Texas



Scale Up Nationally



Total Remaining Across All Data Sources

35,150 WWTPs

Scale our understanding on California and Texas to all remaining states

Spot the difference

Spot the difference

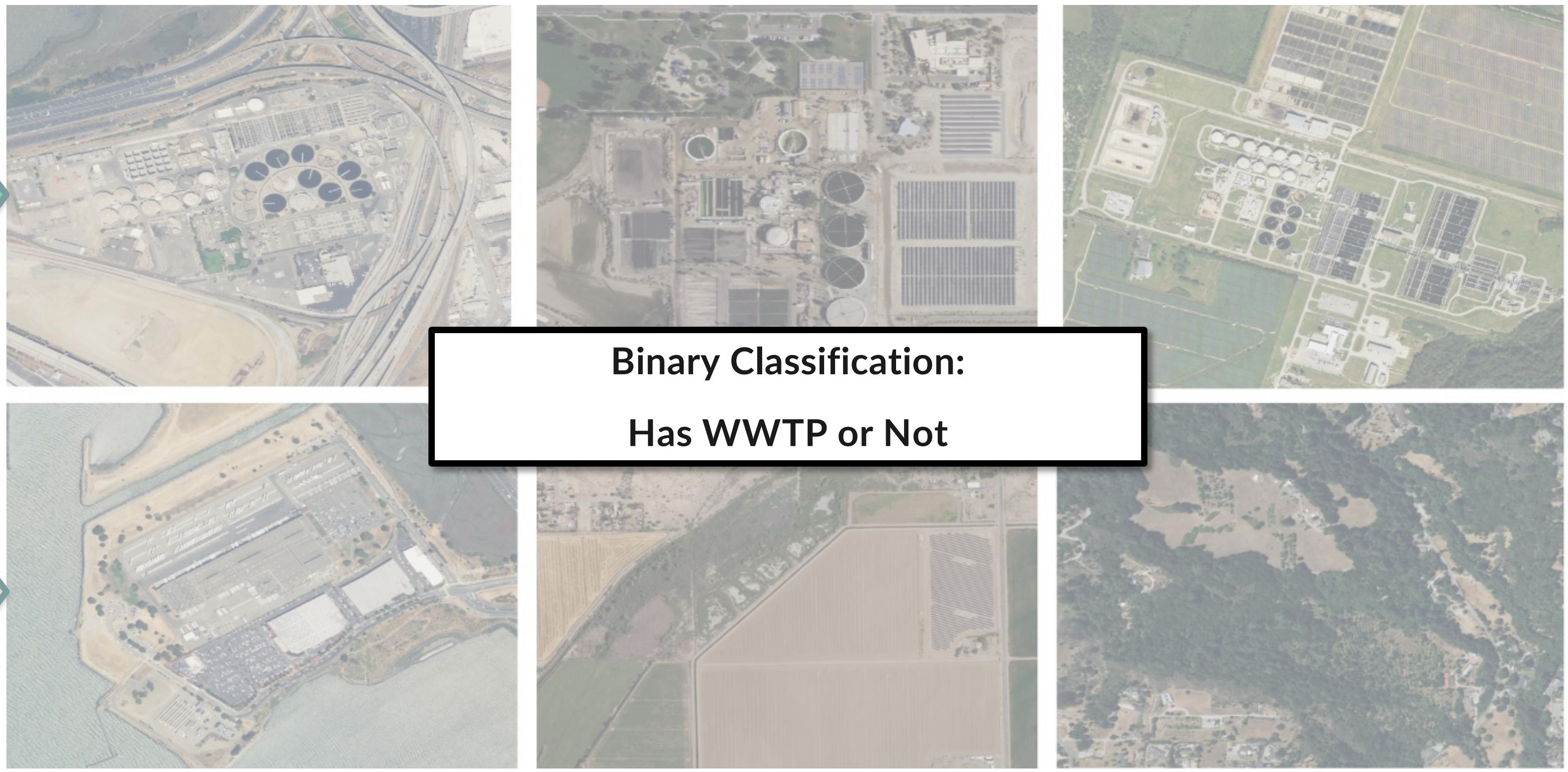
WWTP

NO WWTP

Spot the difference

WWTP

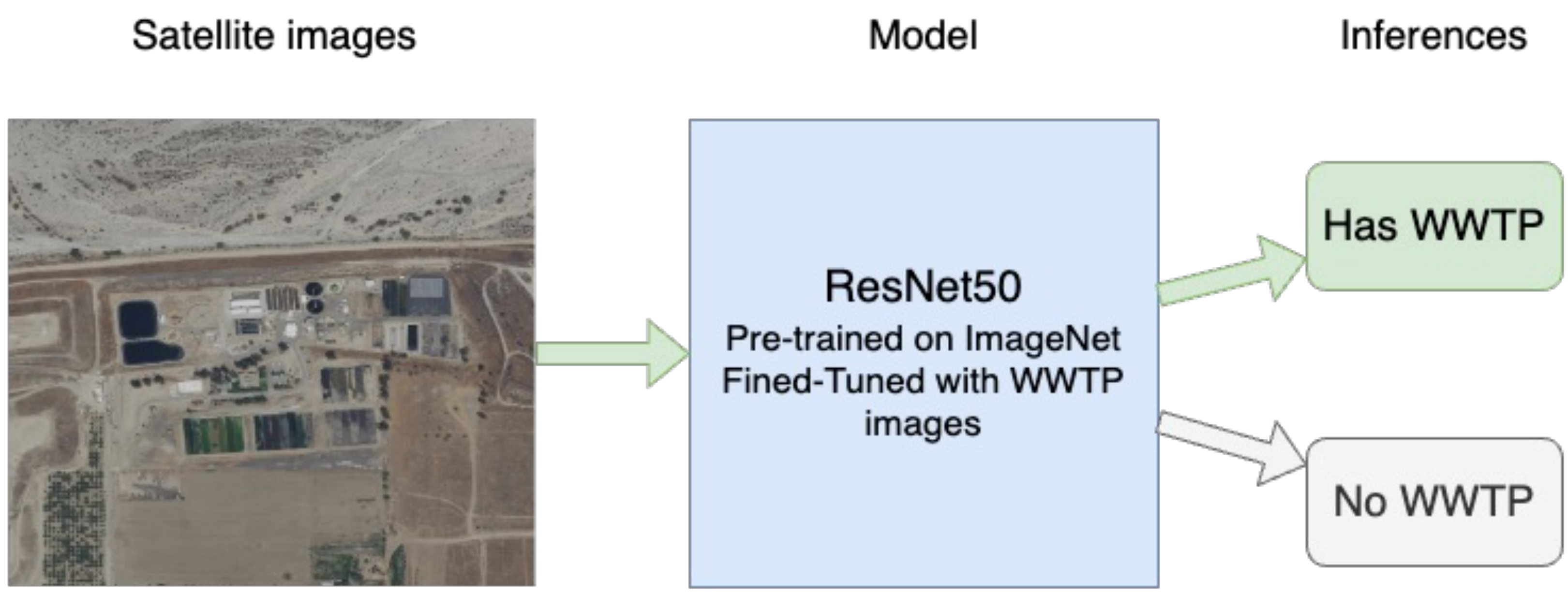
NO WWTP



Binary Classification:

Has WWTP or Not

Scene Classification Model Pipeline



How do we know if the model would perform well with our data and whether it would be the same when scaling up?

Experiment Methodology

Stage 1: How good the model is?

Compare model performance using different crop sizes

Stage 2: Are you sure it would have decent performance when scaling up?

Within domain vs Cross domain performance check

Experiment Methodology

Stage 1: How good the model is?

Compare model performance using different crop sizes

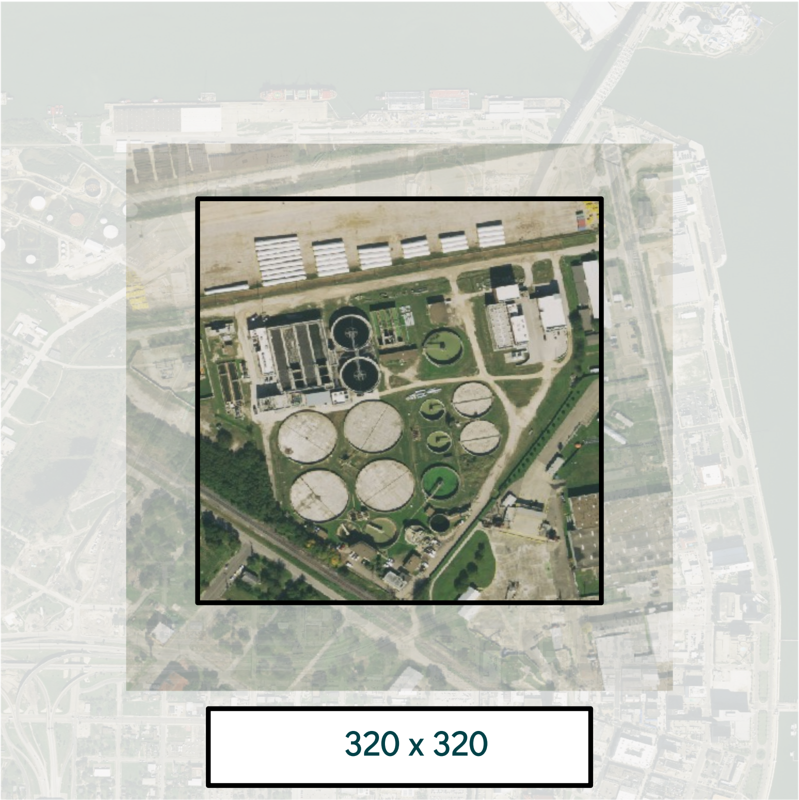
Stage 2: Are you sure it would have decent performance when scaling up?

Within domain vs Cross domain performance check

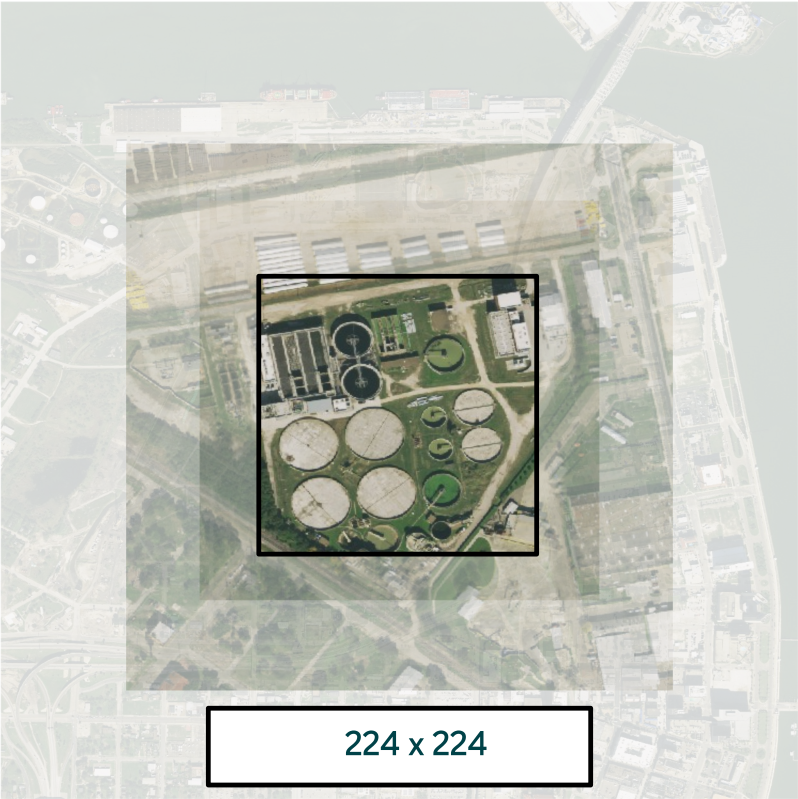
Experiment Stage 1 – Change Crop Size

Experiment Stage 1 – Change Crop Size

Experiment Stage 1 – Change Crop Size



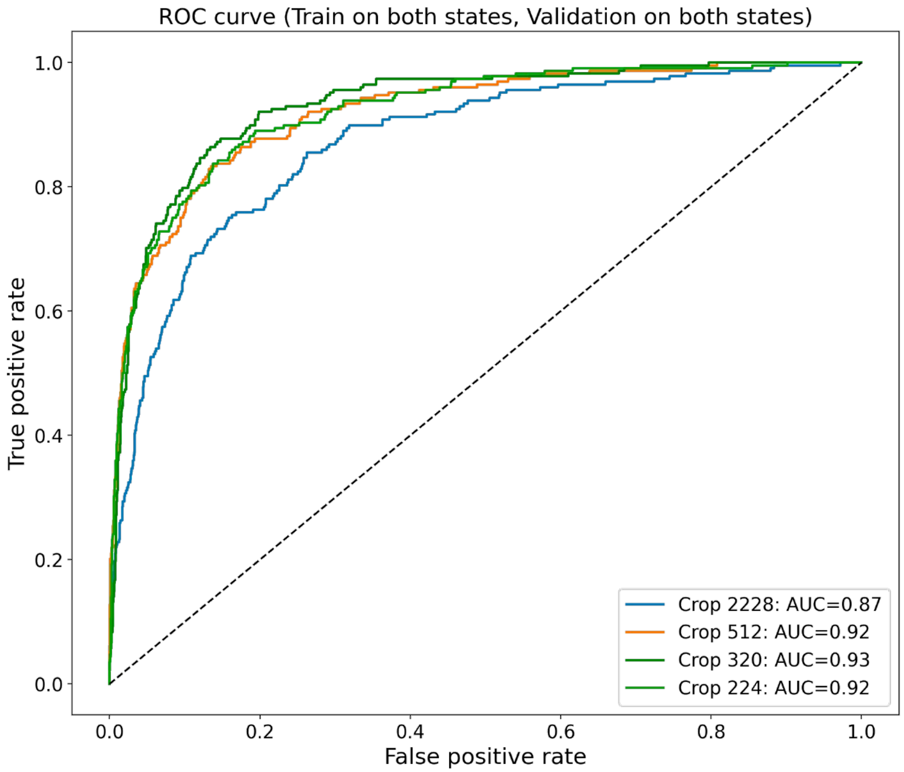
Experiment Stage 1 – Change Crop Size



Experiment Stage 1 – Change Crop Size

Best Model Performance Achieved with Crop Size of 320 x 320

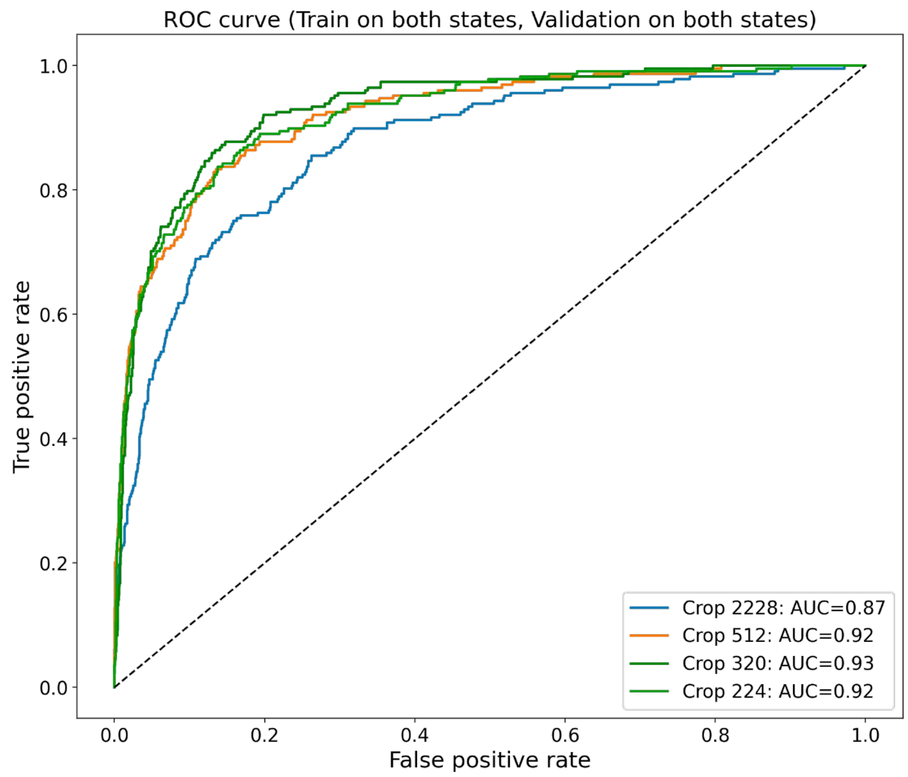
Center Crop Size	AUC	Max F1 score
Original (2228 x 2228)	0.87	0.5198
512 x 512	0.92	0.5075
320 x 320	**0.93**	**0.6554**
224 x 224	0.92	0.5635



Experiment Stage 1 – Change Crop Size

Best Model Performance Achieved with Crop Size of 320 x 320

Center Crop Size	AUC	Max F1 score
Original (2228 x 2228)	0.87	0.5198
512 x 512	0.92	0.5075
320 x 320	**0.93**	**0.6554**
224 x 224	0.92	0.5635



Experiment Methodology

Stage 1: How good the model is?

The model DOES perform well on our data

Stage 2: Are you sure it would have decent performance when scaling up?

Within domain vs Cross domain performance check

Experiment Methodology

Stage 1: How good the model is?

The model DOES perform well on our data

Stage 2: Are you sure it would have decent performance when scaling up?

Within domain vs Cross domain performance check

Experiment Stage 2

Using the best model from Experiment Design (1): Pretrained ResNet50 on ImageNet Weights, Crop size: 320 x 320

		Validation On	
		Texas	California
Training On	Texas	AUC:0.93, Max F1: 0.8454	
	California		AUC:0.91, Max F1: 0.6139

Within Domain: Train in the **same** state as the validation images

- Training with images in the same state as the validation images (Within Domain) yields good performance

Experiment Stage 2

Using the best model from Experiment Design (1): Pretrained ResNet50 on ImageNet Weights, Crop size: 320 x 320

		Validation On	
		Texas	California
Training On	Texas	AUC:0.93, Max F1: 0.8454	AUC:0.89, Max F1: 0.5482
	California	AUC:0.75, Max F1: 0.6175	AUC:0.91, Max F1: 0.6139

Within Domain: Train in the **same** state as the validation images
Cross Domain: Train in a **different** state from the validation images

- Despite a slight drop from within-domain results, Cross Domain training still presents good max F1 scores and AUC metrics, which demonstrates our model's generalization and predictive ability across disparate WWTP sites.

Experiment Methodology

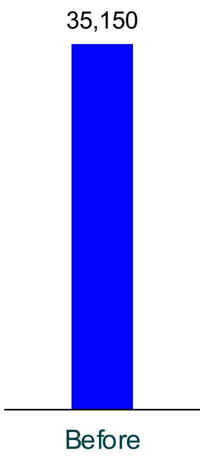
Stage 1: How good the model is?

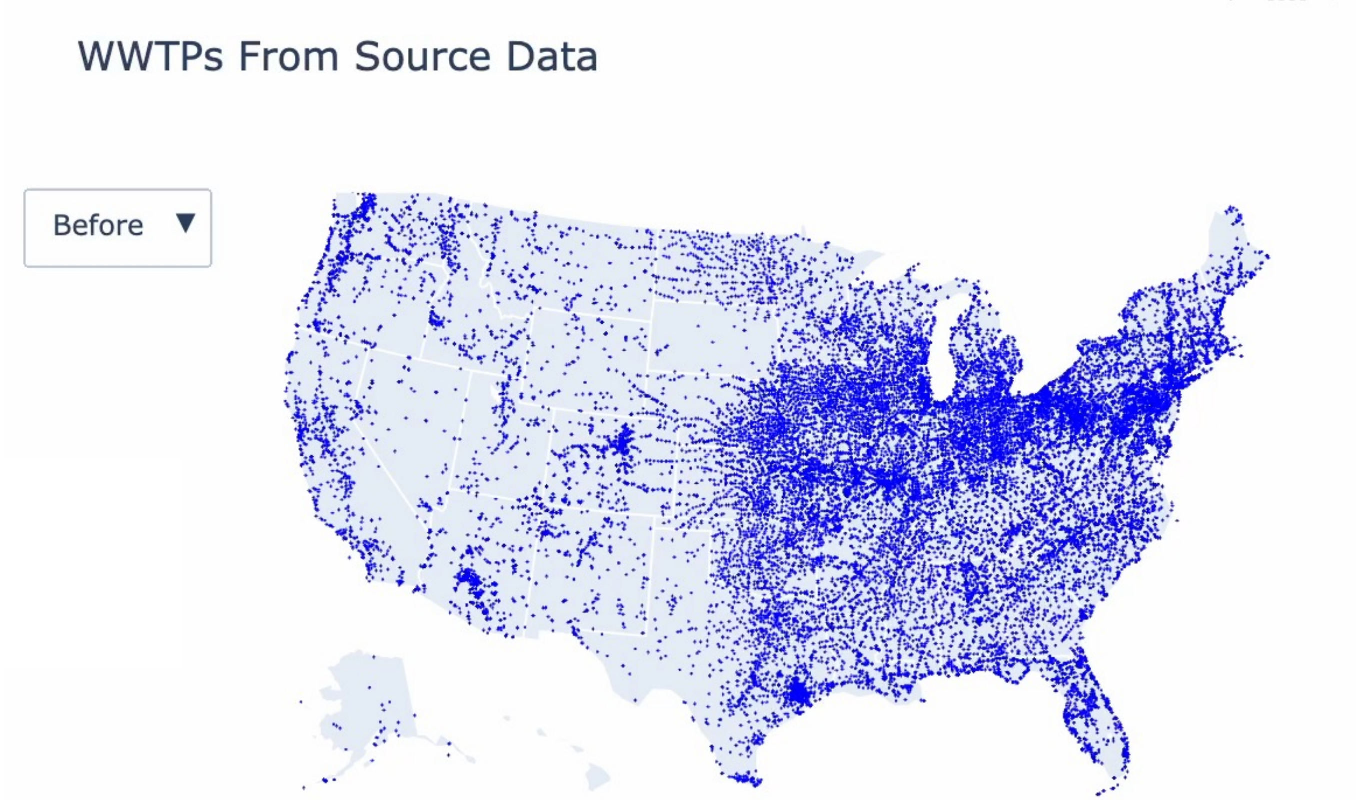
The model DOES perform well on our data

Stage 2: Are you sure it would have decent performance when scaling up?

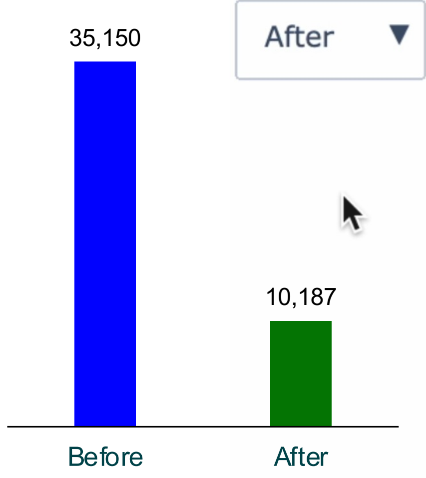
The model WOULD perform well on other states

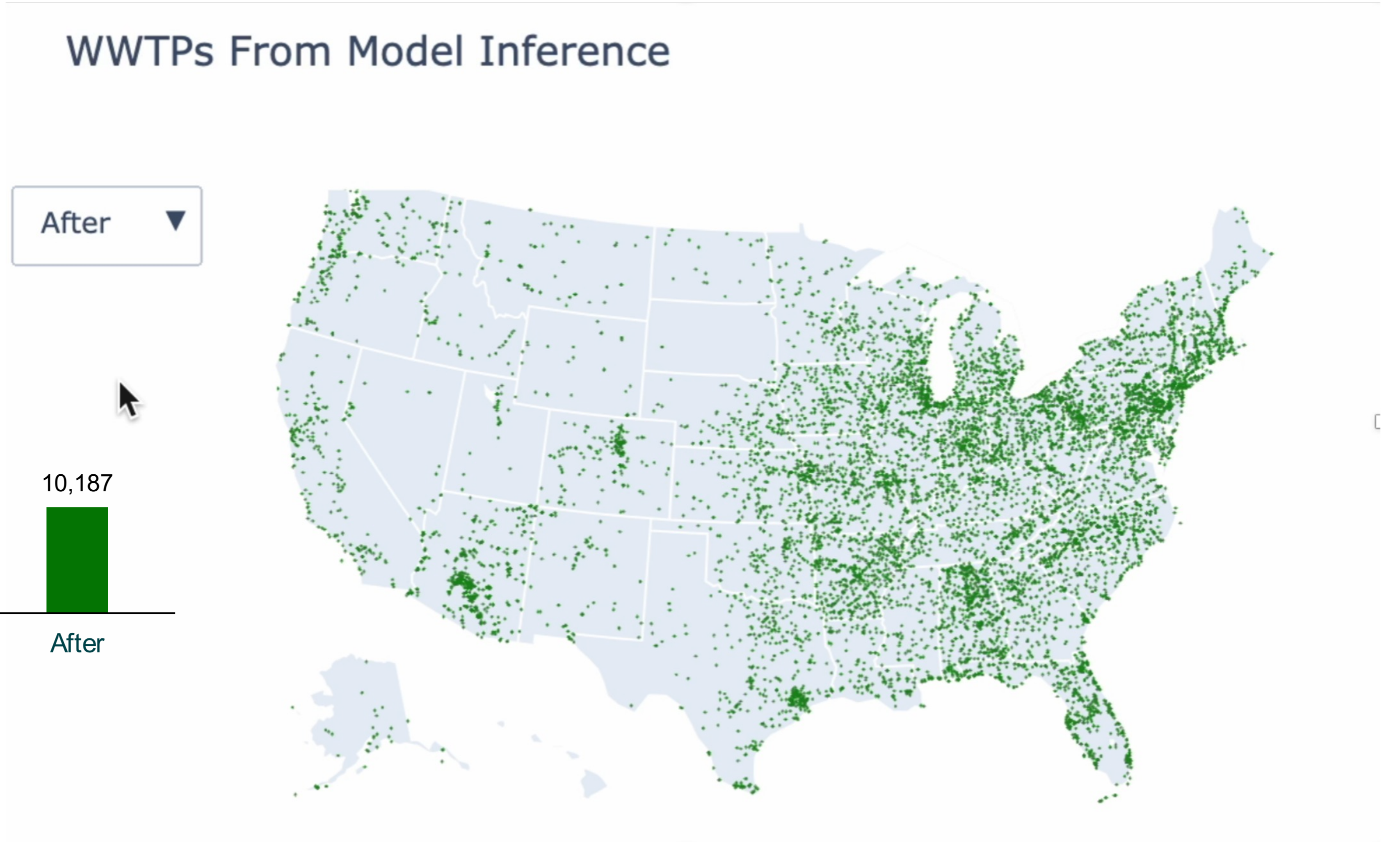
Inference Results



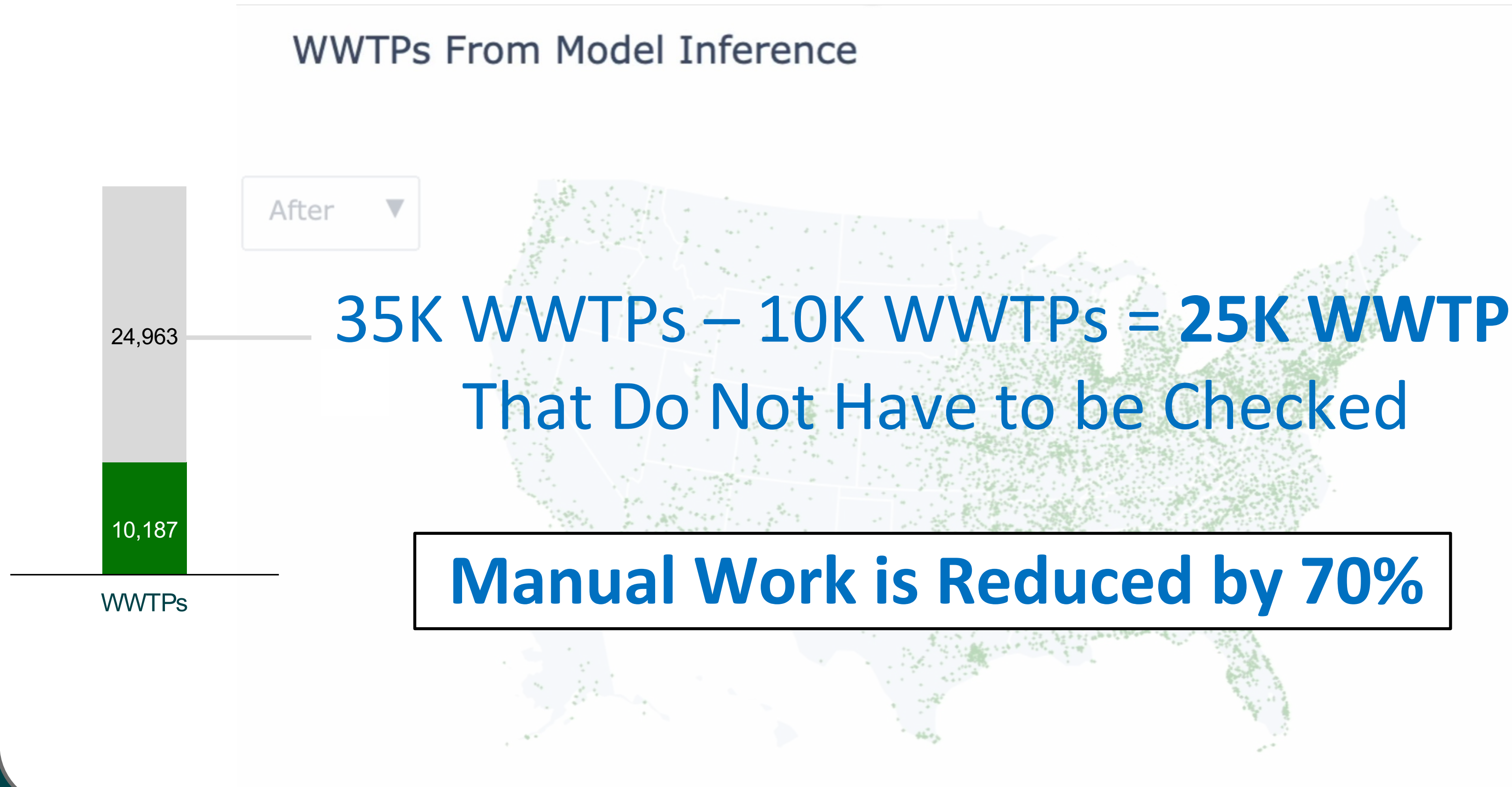


Inference Results





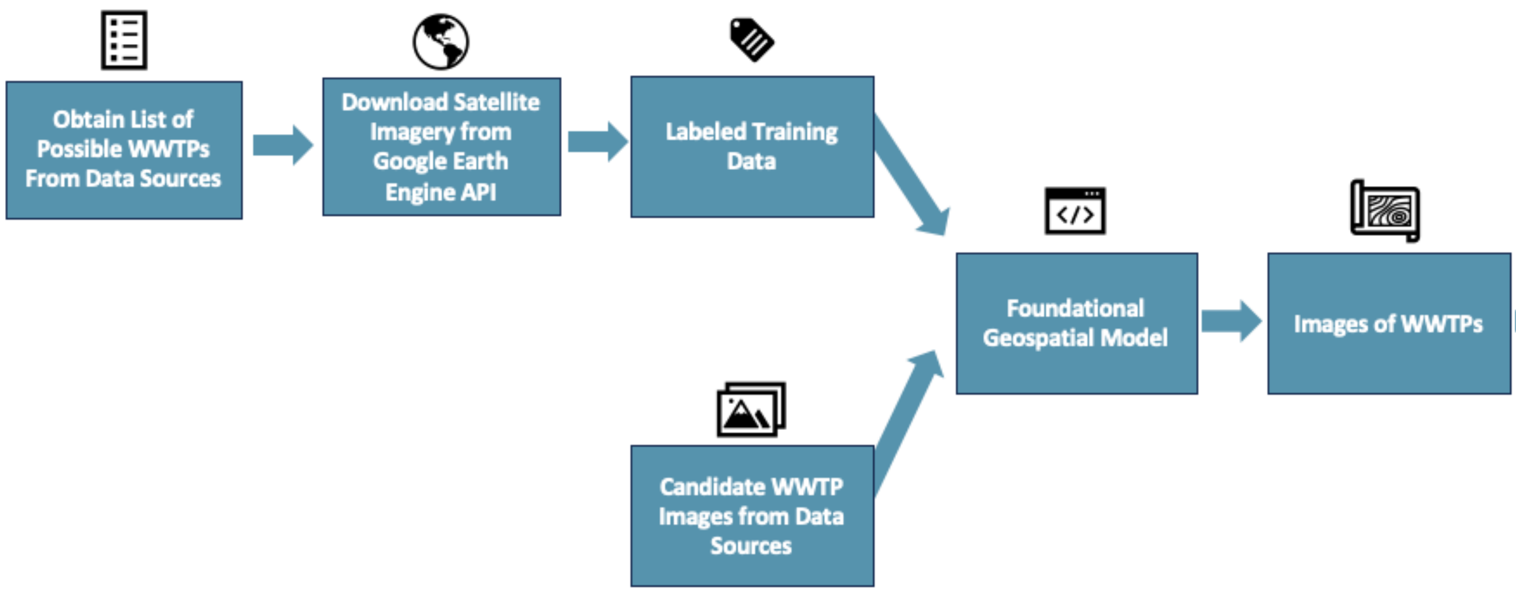
Inference Results



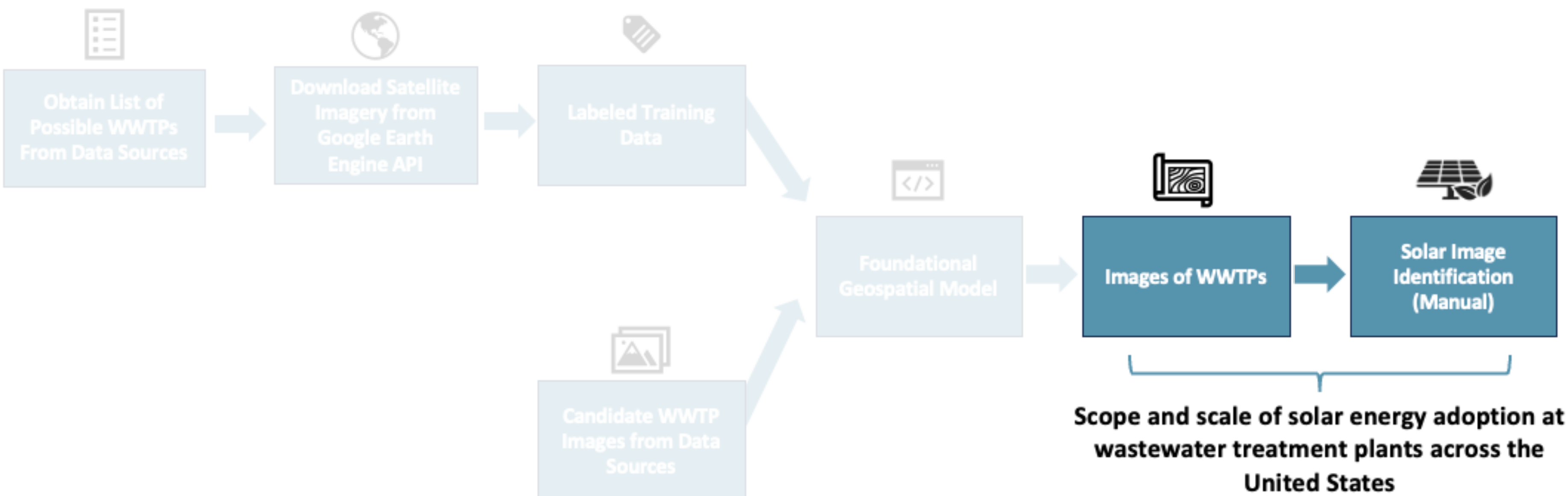
35K WWTPs – 10K WWTPs = **25K WWTPs**
That Do Not Have to be Checked

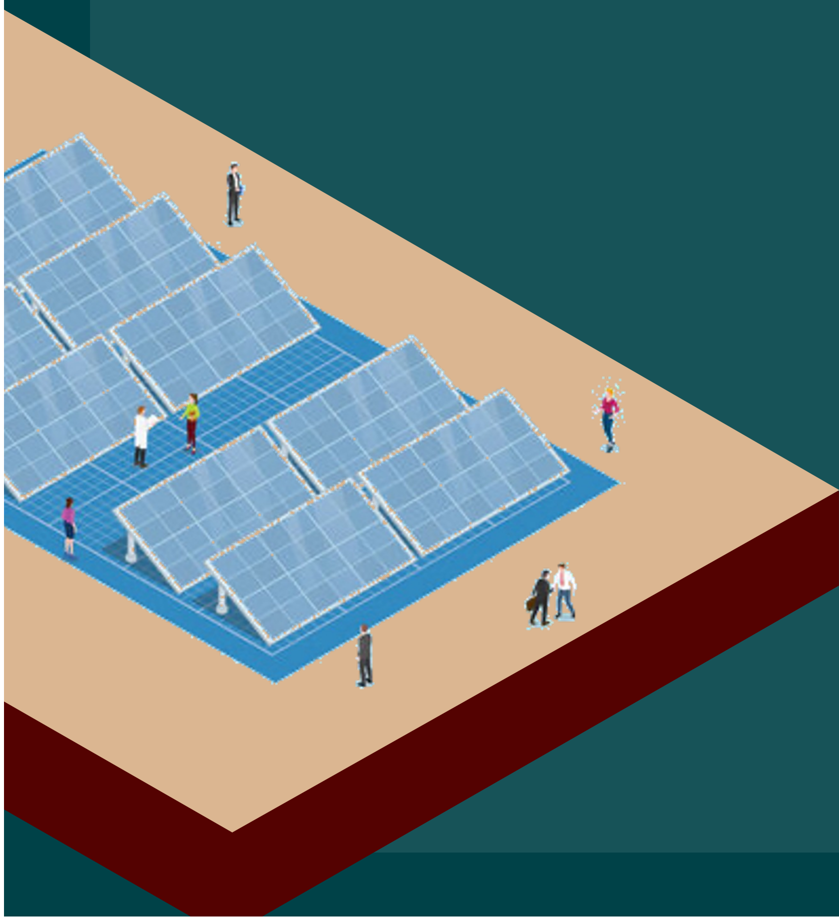
Manual Work is Reduced by 70%

Project Approach



Project Approach





Client Impact & Value Added


Identify patterns in solar adoption at Wastewater Treatment Plants

Observations

Solar adoption is **14 percentage points greater in California than Texas**

Possible Future Analyses

Does WWTP size and capacity play a role in solar adoption?

Does solar adoption depend on state?

If so, why and factors affect this?


Do surrounding communities benefit from solar adoption?

Observations

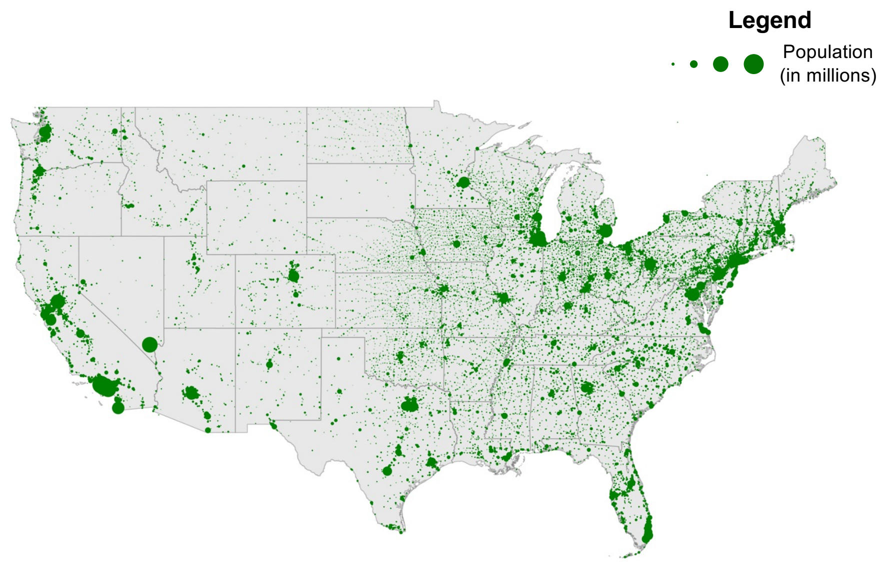
258 million people live surrounding WWTPs

Combining with socio-economic and wastewater service charge data

Possible Future Analyses

Does solar adoption lower wastewater service cost?

How does solar adoption benefit surrounding communities economically?



THANK YOU!

Special Thank You To Our Advisors:

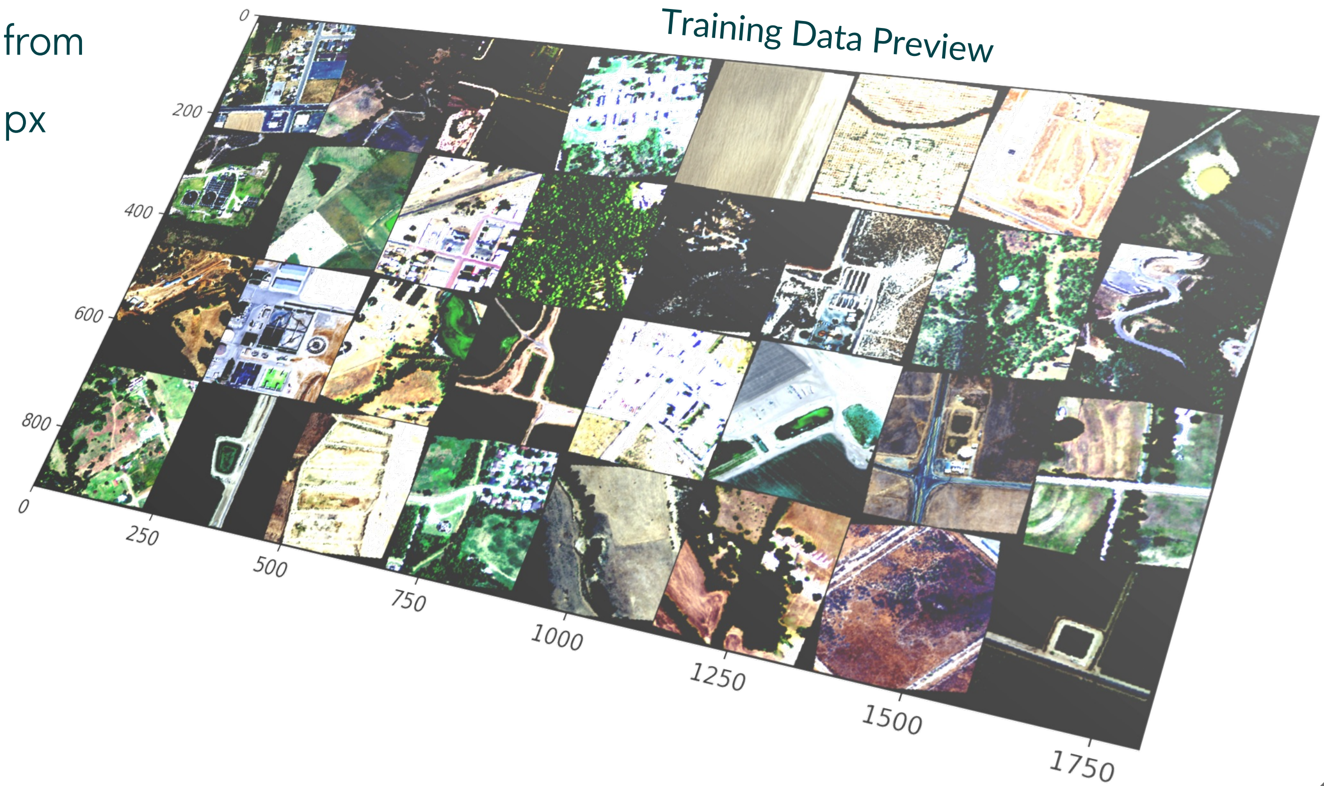
Christine Kirchhoff
Ph.D, P.E.
Penn State University

Kyle Bradbury
Ph.D.
Duke University

APPENDIX

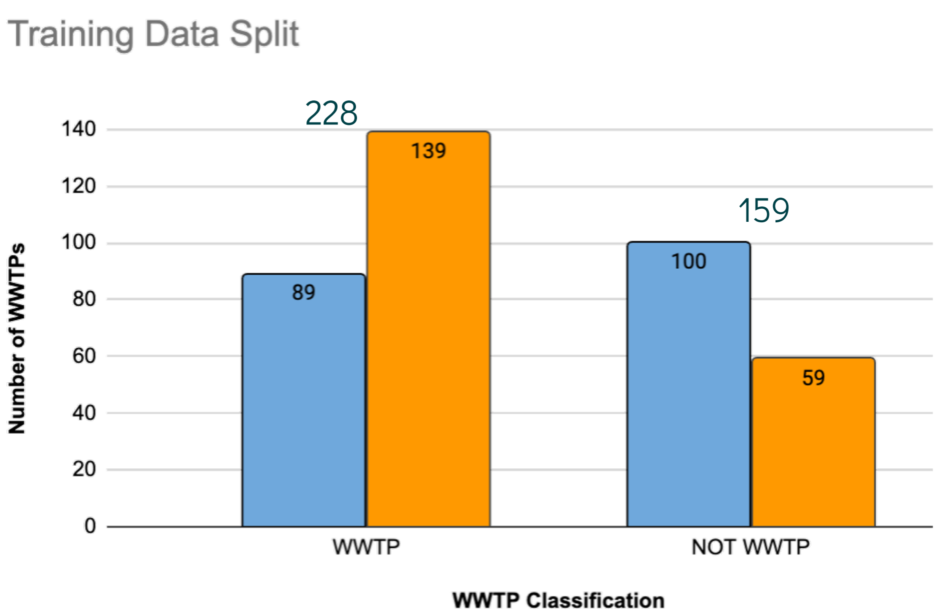
Data Pre-processing

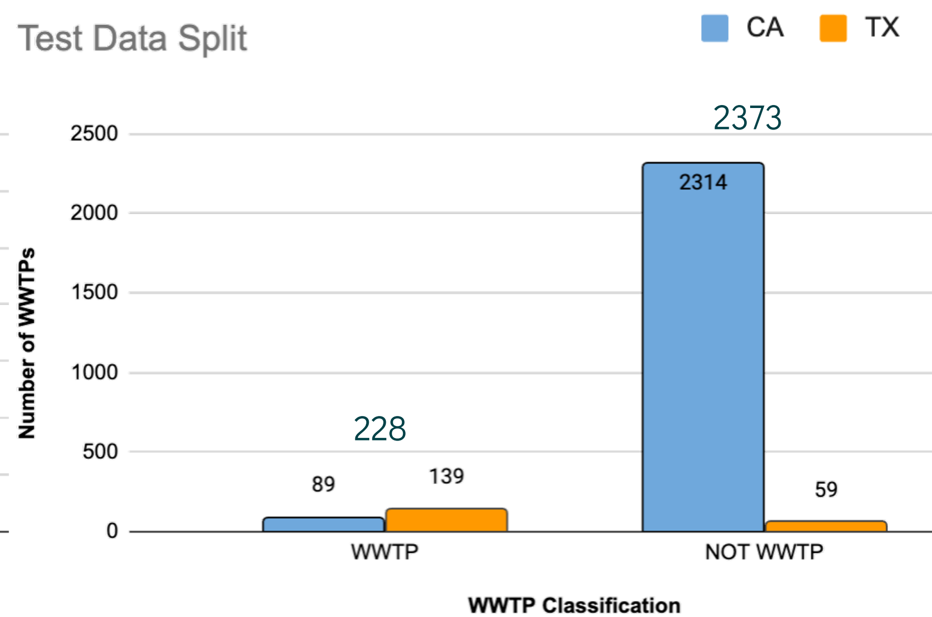
- Center crop 320x320 px from the original 2228 x 2228 px
- Resize to 224 x 224 px
- Convert to tensor
- Normalize
  - μ = [0.485, 0.456, 0.406]
  - σ = [0.229, 0.224, 0.225]
- Random flip
- Random rotation



Data preparation for Experiment

Randomized Split to achieve “balanced” data





Model Architecture