

Predicting Trump and Sanders Primary Victories

Alan Schoen

Data: Demographics and
Primaries

Dataset



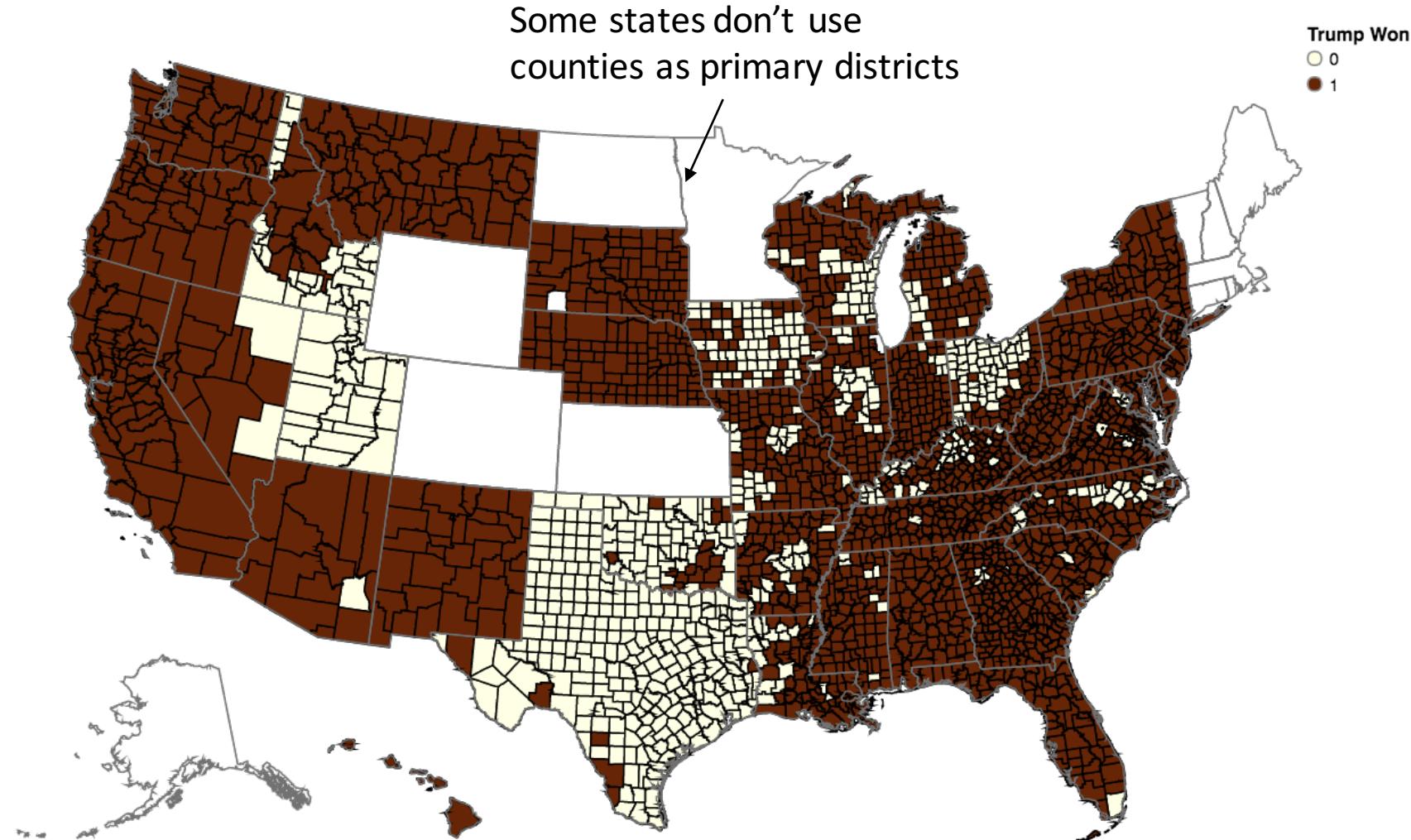
ELECTION
CENTRAL

kaggle

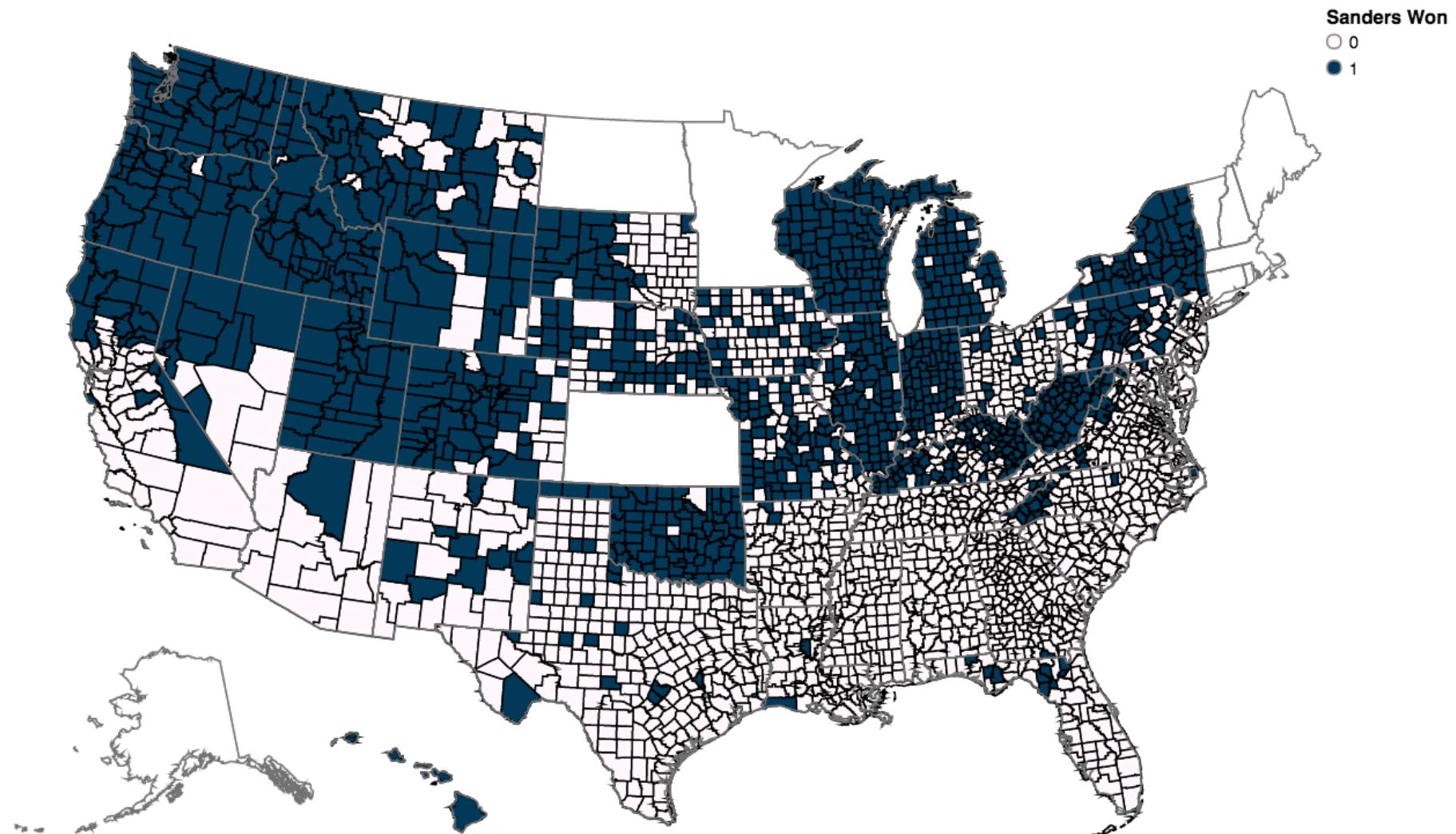


- All predictors were available in 2014
- Demographics + Primary results in 2,710 counties
- 57 total features(59 after primary type is expanded to 3 booleans)
- Demographics: race, age, population, domiciles
- Economics: Unemployment, jobs in various industries
- Primaries: date, format

Target 1: Donald John Trump Primary results



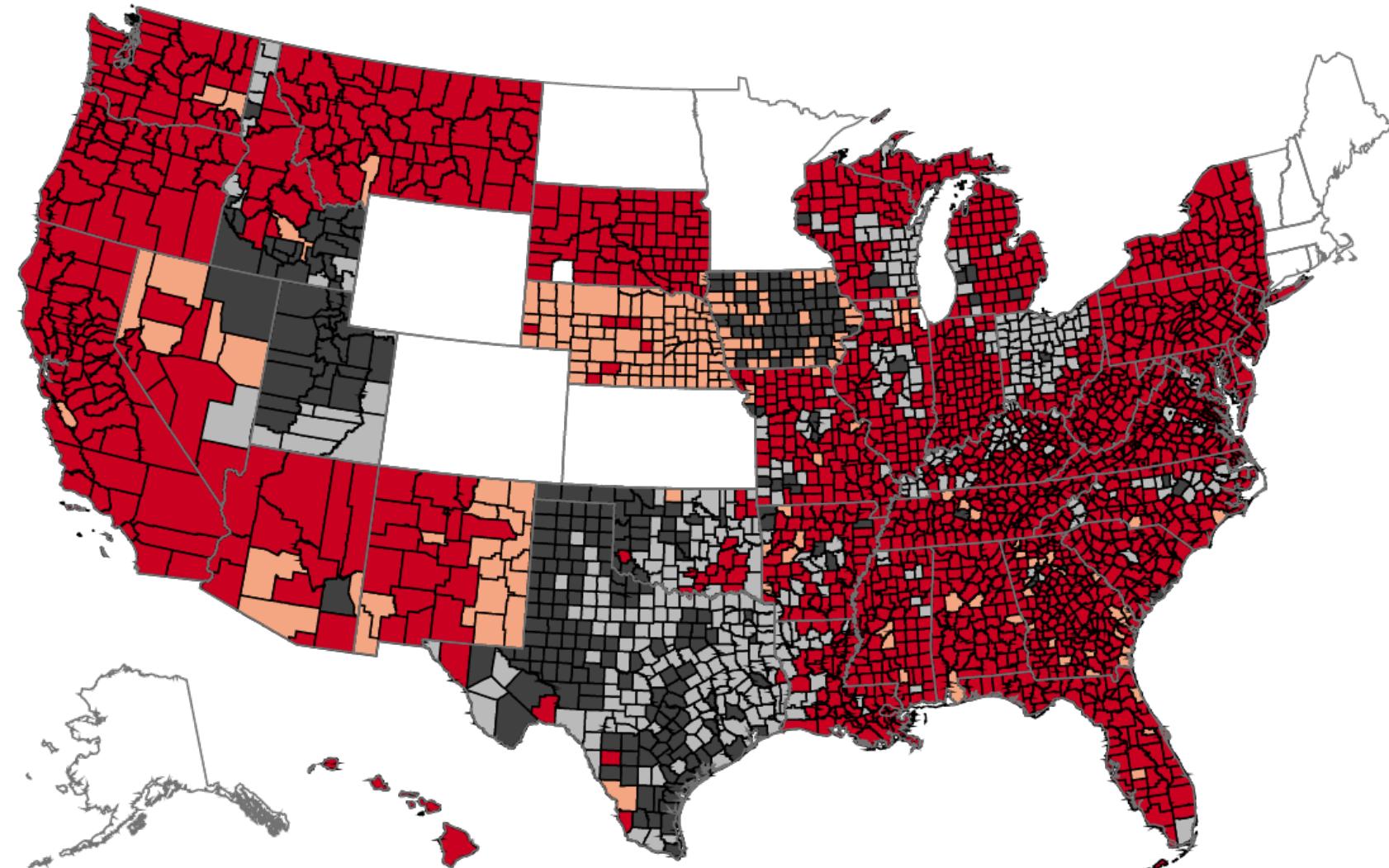
Target 2: Bernard Sanders Primary Results



Results

Model Results: Trump

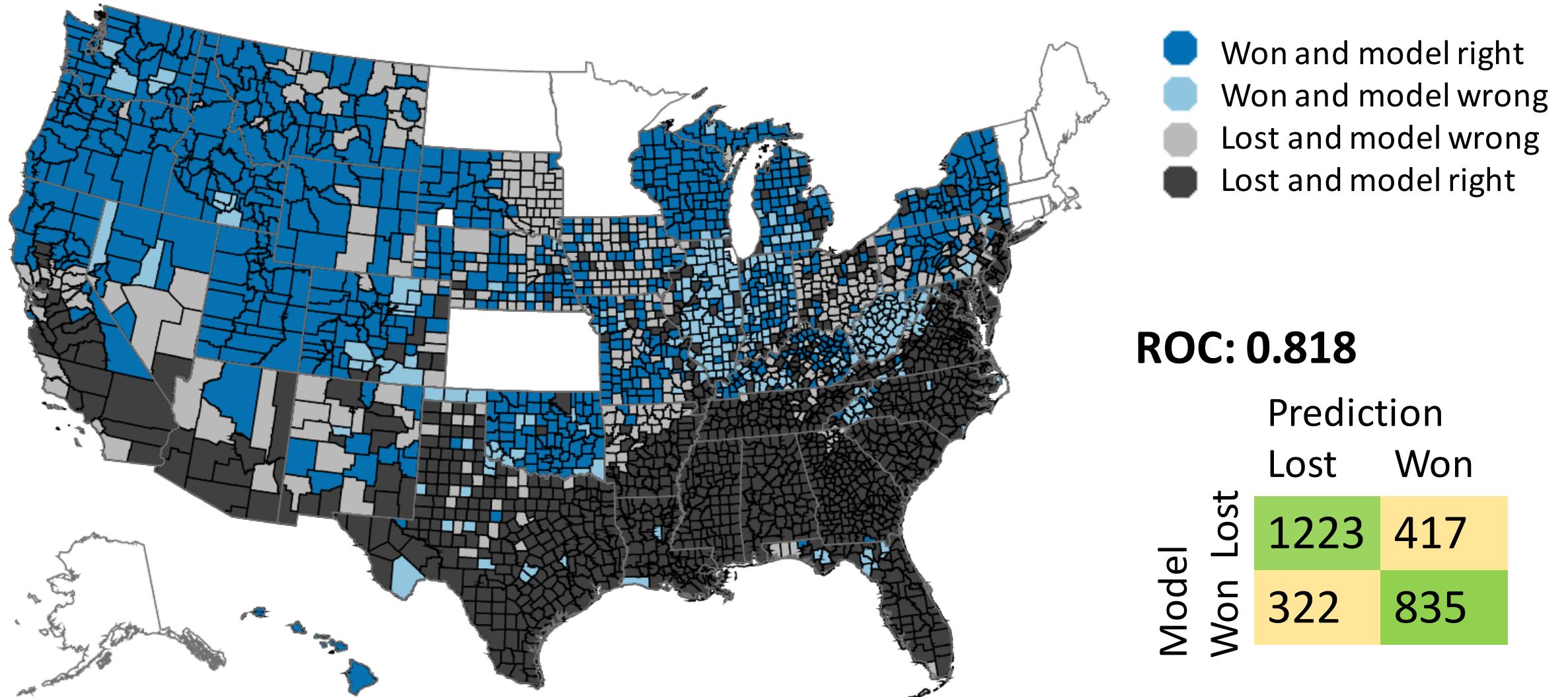
5-fold cross-validation.
Test set results shown.



ROC: 0.765

		Prediction	
		Lost	Won
Model	Won	285	428
	Lost	210	1787

Model Results: Sanders





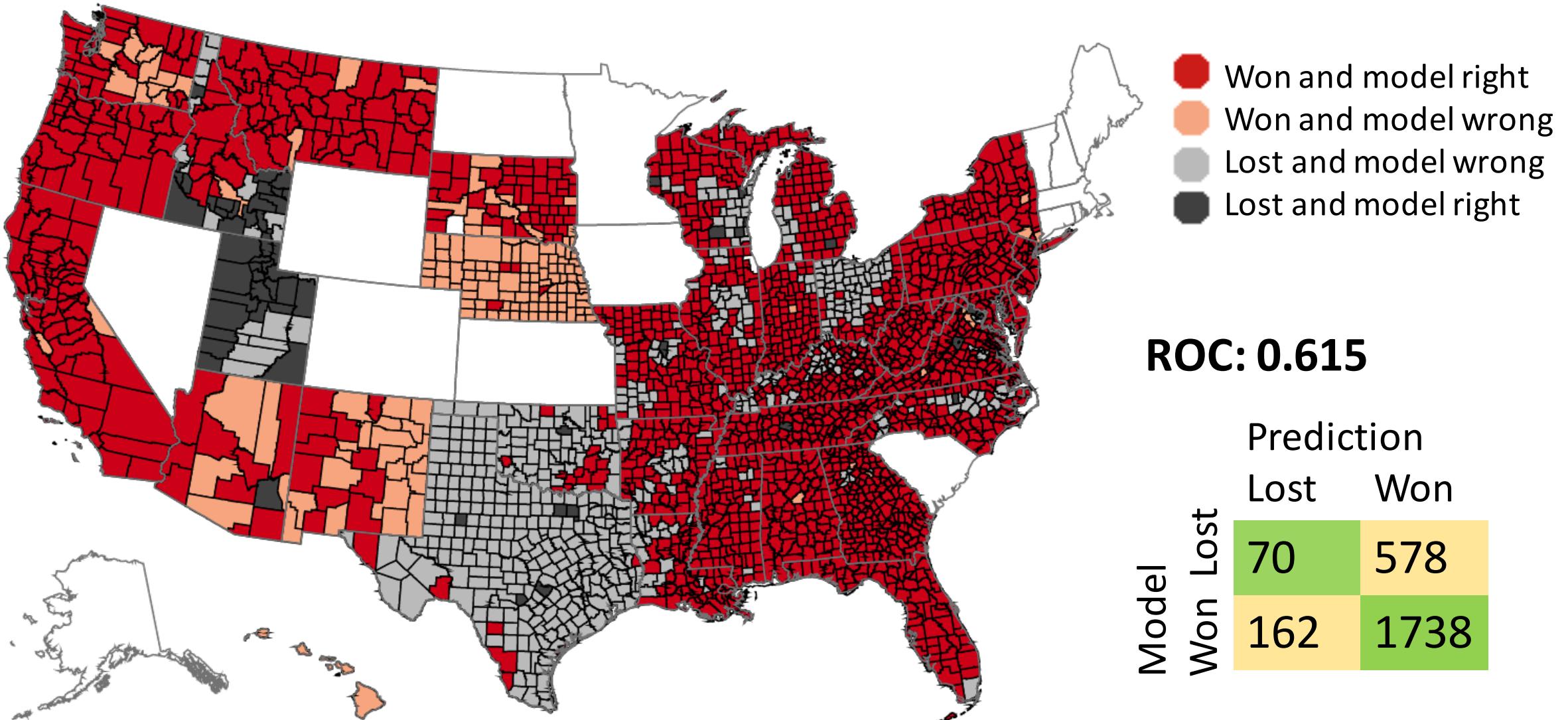
A photograph of a man with short brown hair, wearing a dark suit jacket, a white shirt, and a red patterned tie. He is looking downwards with a neutral to slightly weary expression. The background is a blurred indoor setting with blue and white tones.

CHUMP!

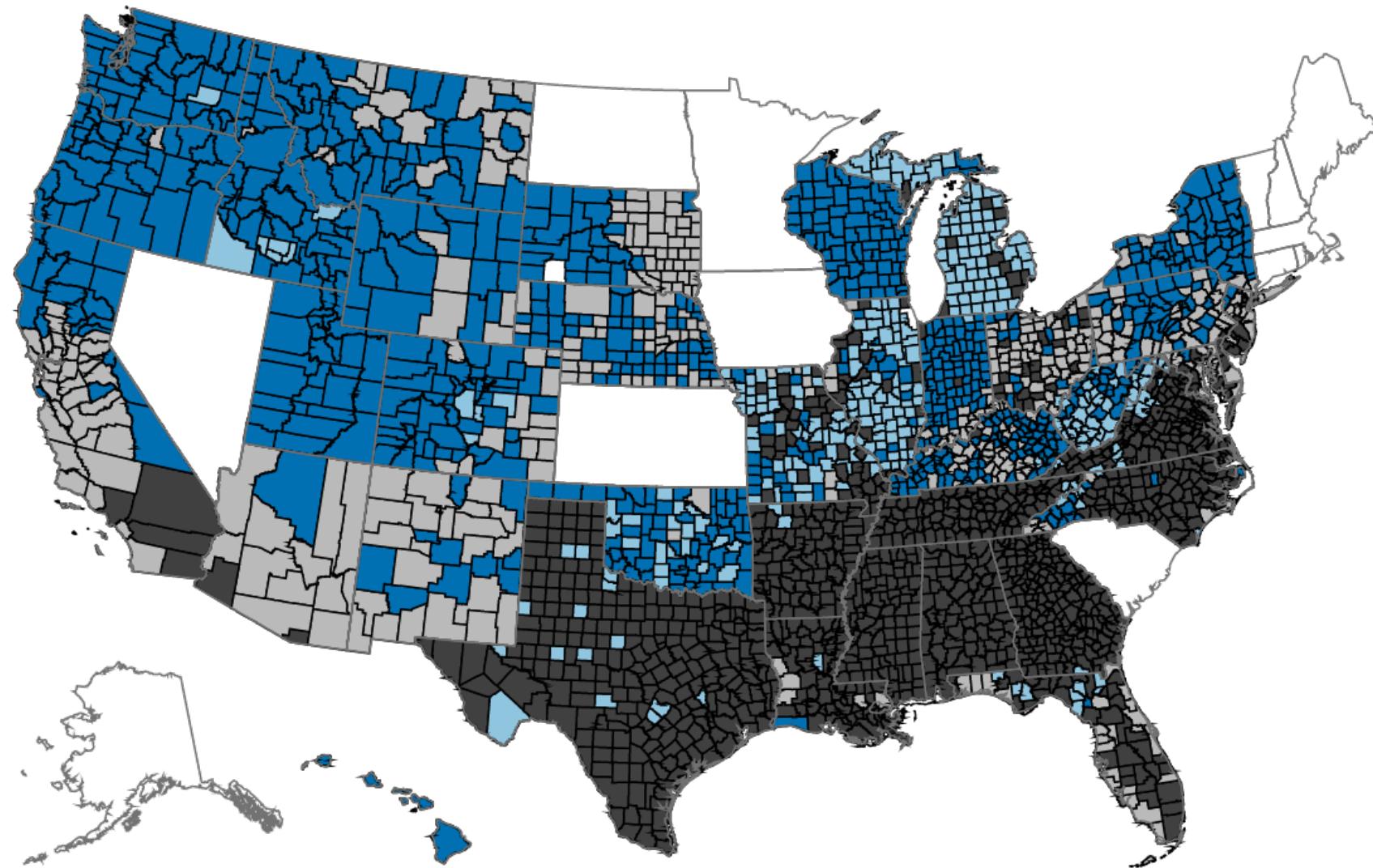
Not exactly. I cheated!

- Training/test split was random so the model included hindsight
- What if I train primary-by-primary and evaluate results?
- Next model: Starting on Super Tuesday (March 1) predict each primary from the primaries before

Rolling Model Results: Trump



Rolling Model Results: Sanders



- Dark Blue: Won and model right
- Light Blue: Won and model wrong
- Gray: Lost and model wrong
- Black: Lost and model right

ROC: 0.819

		Prediction	
		Lost	Won
Model	Won	1120	409
	Lost	302	804



Please Clap