



# **Modeling Voter Turnout**

## **in US House of Representatives Elections**

**Audrey Baker**

**January 26, 2018**

# Agenda

---

1. Background
2. Exploratory Analysis
3. Modeling
4. Insights and Next Steps

# Background

---

# Objective

---

Predict votes cast in each congressional district for a US House of Representatives Race in order to help national political groups determine optimal allocation of resources for canvassing and 'Get Out The Vote' operations.

# Data sources

---

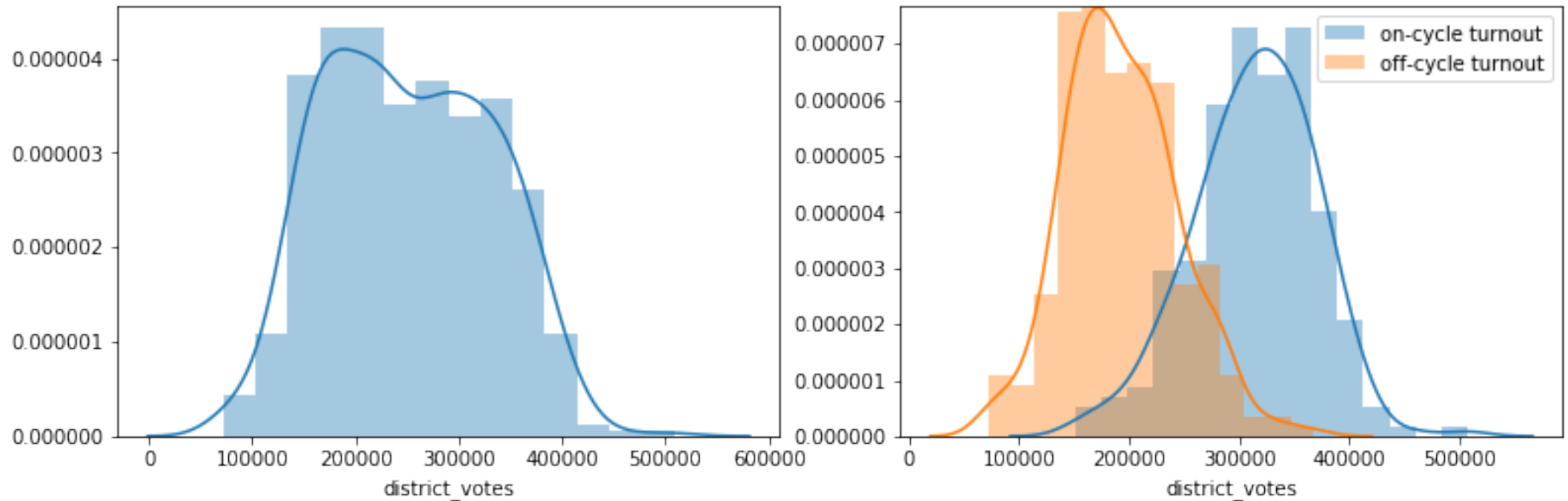
1. Vote counts by district in 2014 and 2016 elections, from US House of Representatives election statistics records
2. Dollars raised and spent by each candidate for election in 2014 and 2016 elections, from Federal Election Commission campaign finance records

# Exploratory Analysis

---

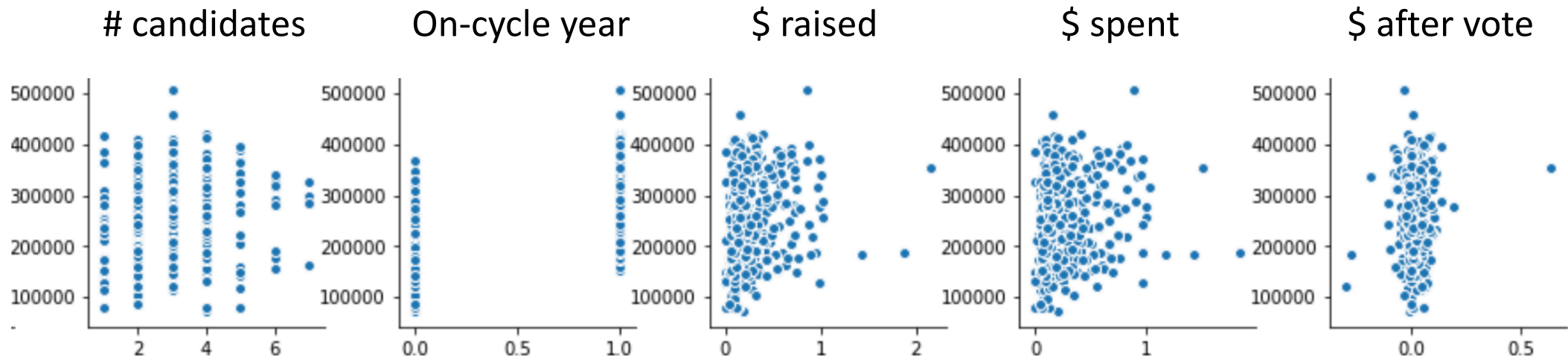
# Distribution of voter turnout

---



# Feature correlations with voter turnout

---

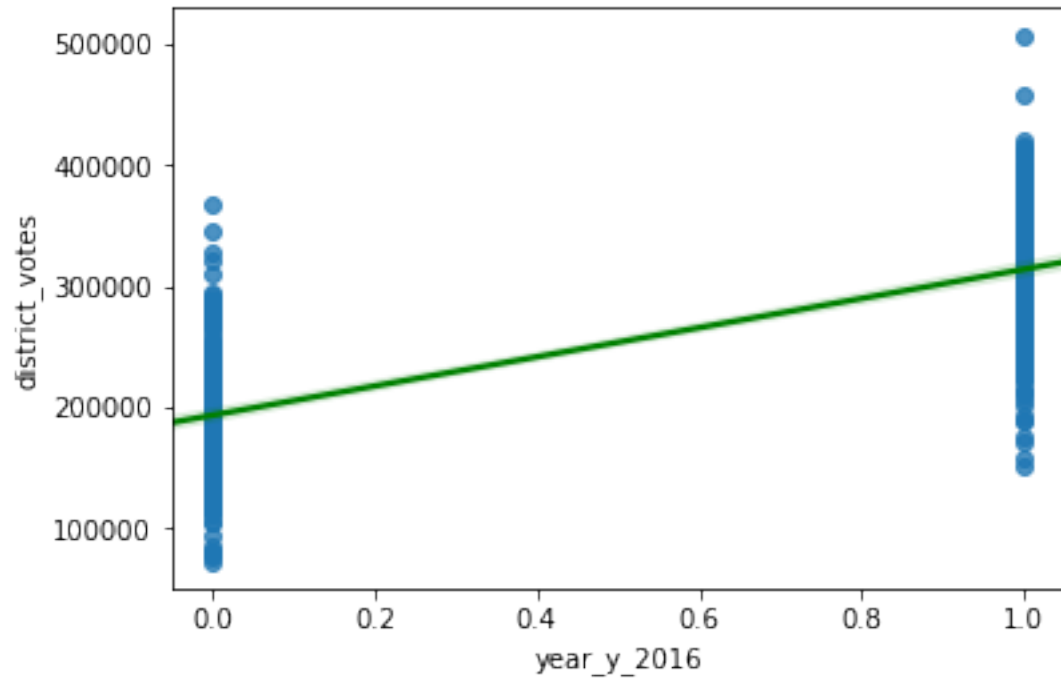




# Modeling

---

# Bivariate model, on-cycle year

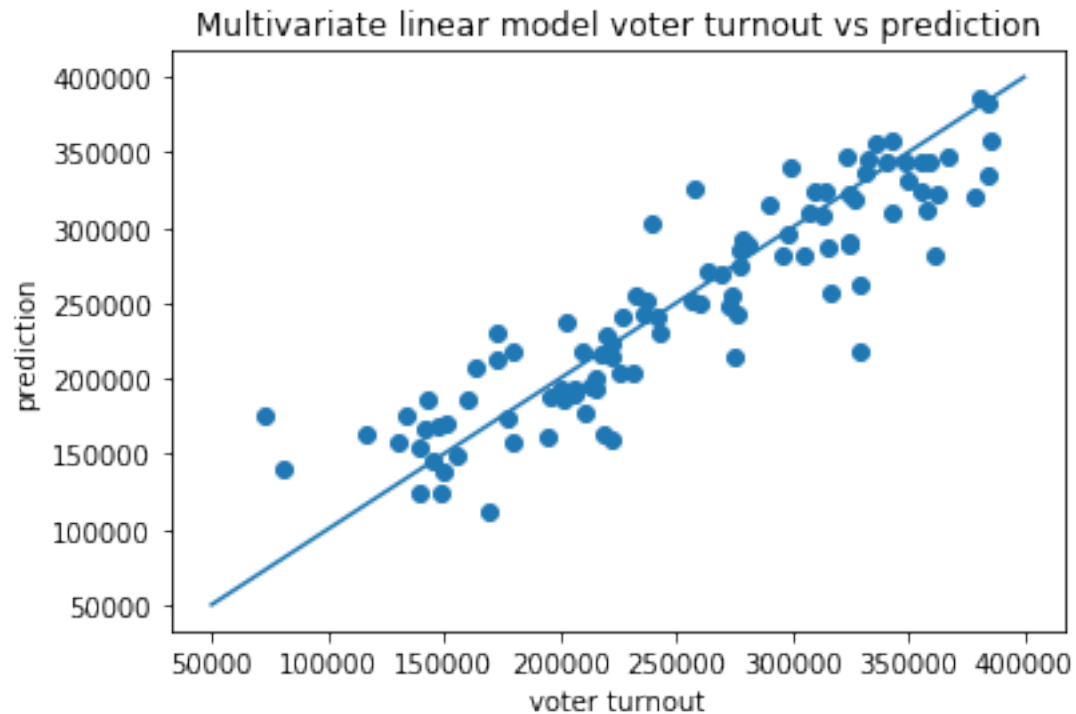


R <sup>2</sup>	
Training data	0.5420345
Testing data	0.6586412

## Model

Votes cast =  $1.9\text{e}4 + 1.2\text{e}5(\text{On-cycle}) + \text{error}$

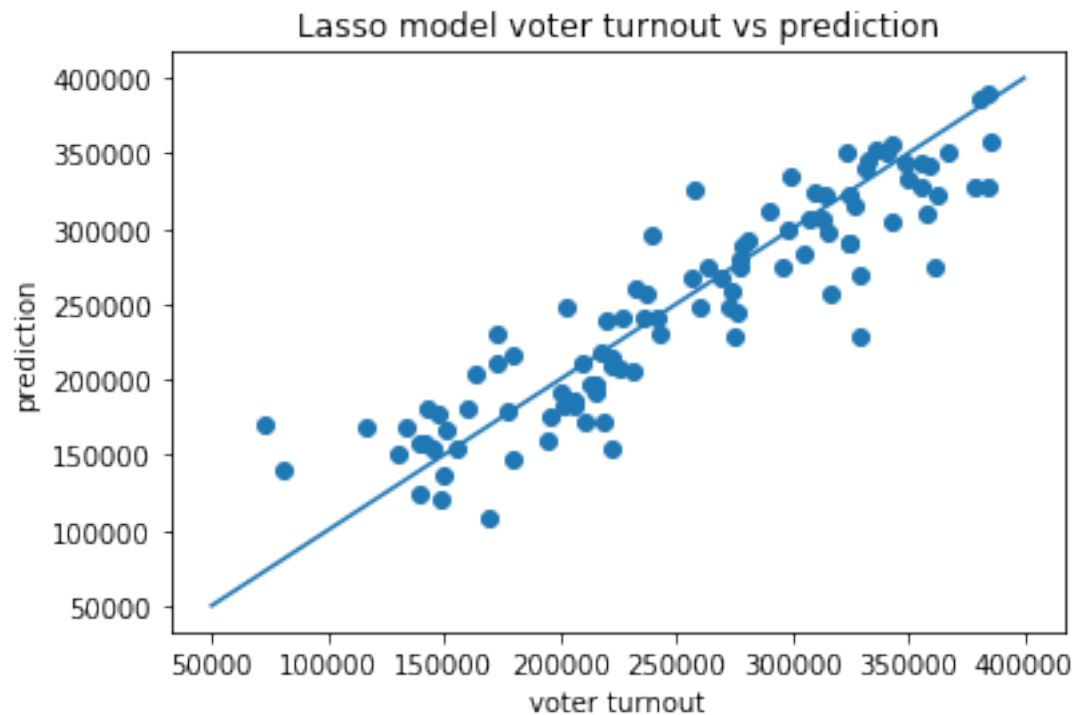
# Multivariate linear regression



R <sup>2</sup>	
Training data	0.8521351
Testing data	0.8141485

Top Features	
\$ at campaign end	-2.12e12
\$ raised	2.12e12
\$ spent	-2.12e12

# Lasso regression



**$R^2$**

Training data

0.8596246

Testing data

0.8187493

**Top Features**

Montana

1.47e5

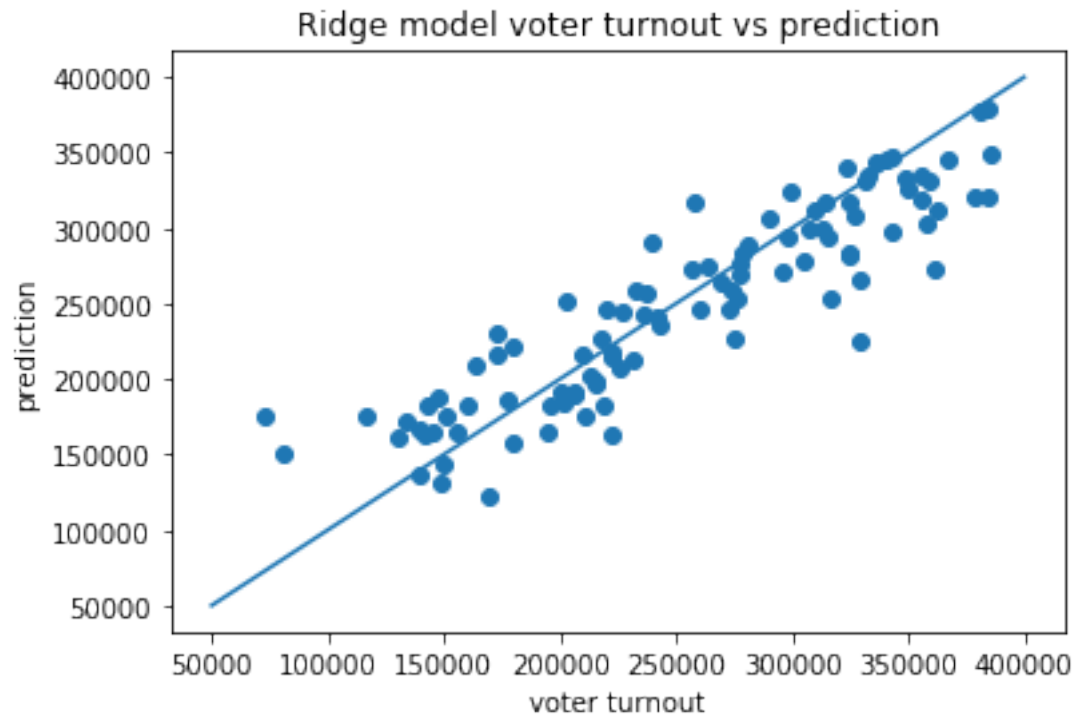
On-cycle

1.17e5

Rhode Island

-8.78e4

# Ridge regression

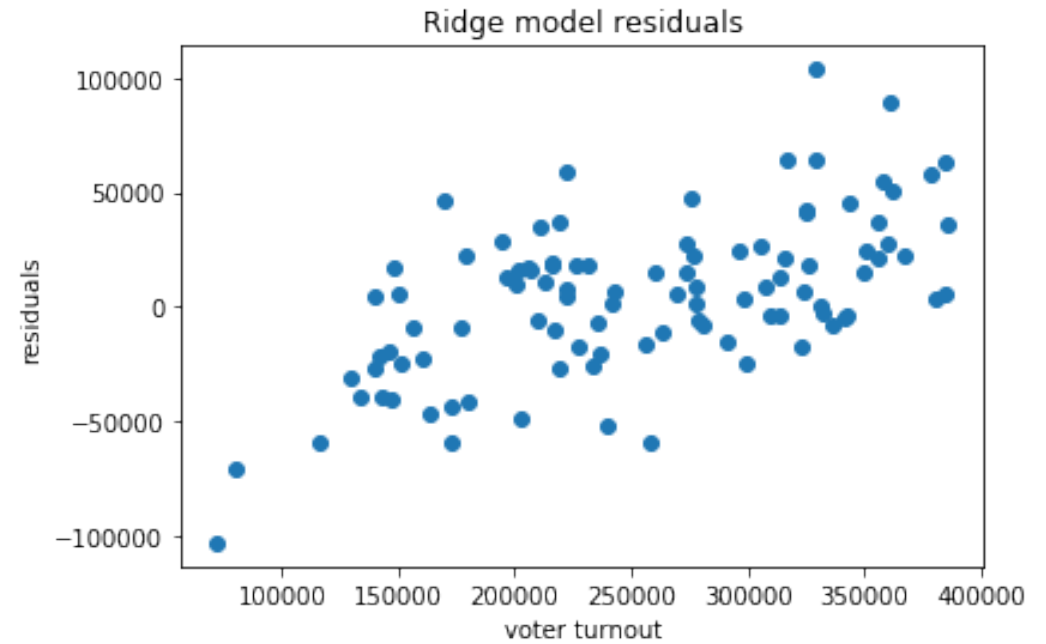
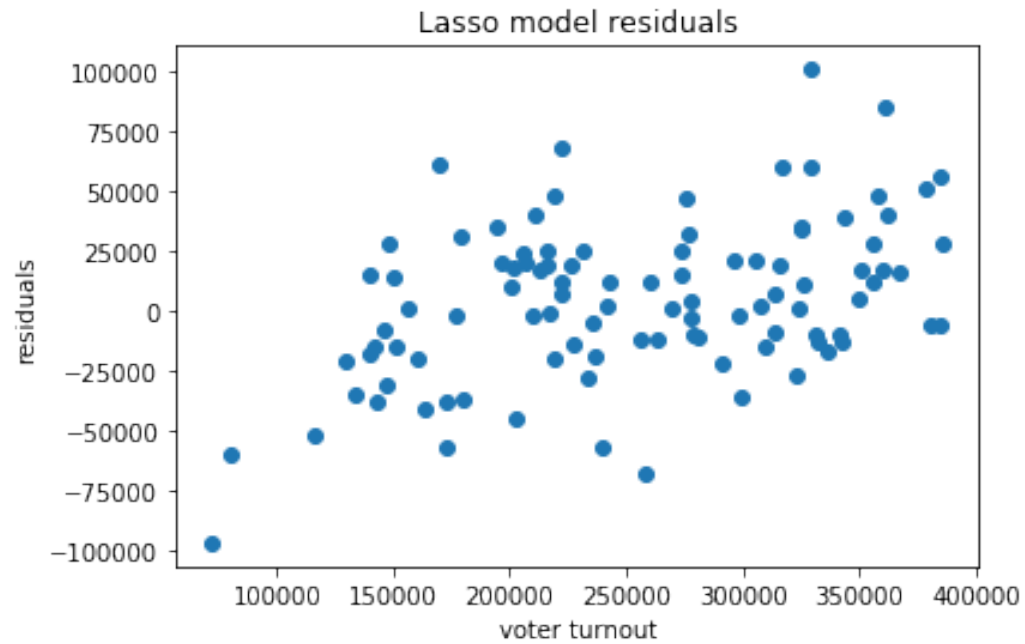


R <sup>2</sup>	
Training data	0.8521260
Testing data	0.8059614

Top Features	
Montana	1.46e5
On-cycle	1.06e5
Maine	8.28e4

# Residuals

---



# Insights and Next Steps

---

# Takeaways

---

1. State average turnout and whether the election is on- or off-cycle are strongest predictors
2. This model is unlikely to beat local, within-state knowledge on resource allocation
3. This model will over-predict low turnout and under-predict high turn-out



# A word of caution

---

Although  $R^2$  is high for unseen data, all models are based on only two election years

- 2016 is widely considered an unusual election
- Midterm elections affect Republican and Democratic areas differently depending on who controls the presidency

# Next steps

---

- Gather data on more election cycles
- Incorporate additional features, e.g. ACS census data from year prior to election, incumbency, political lean, economic trends, presidential and congressional approval ratings