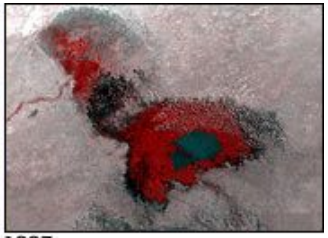
# An analysis of the long-term predictability of drought and its relation to median income

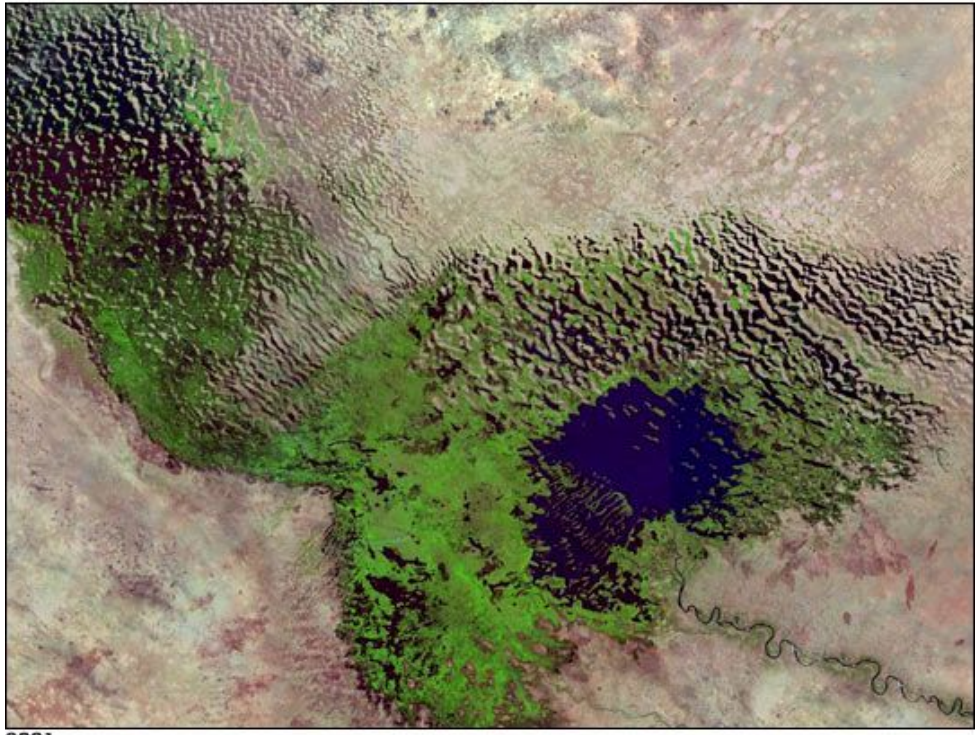
By: Seun Eisape & Victor Nault

1973

1987



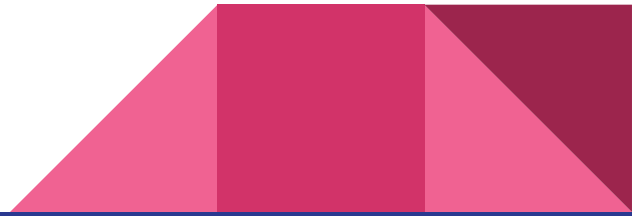
1997



2001

# Our Datasets

- The DSCI for every county in the United States, for every week from January 2nd, 2000 to November 17th, 2021
- The median household income of every county in the United States for the year of 2019



# What is DSCI?

## Example of computing DSCI using categorical USDM statistics


### Percent Area in U.S. Drought Monitor Categories

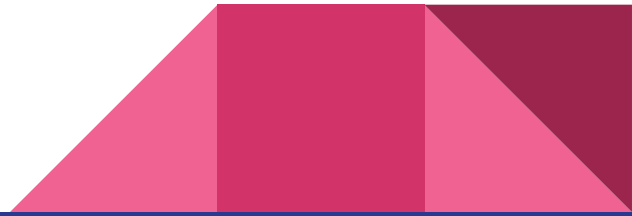
Show 25 entries Search:

| Week | None | D0 | D1 | D2 | D3 | D4 | DSCI |
|---|---|---|---|---|---|---|---|
| 2017-08-29 | 0.41 | 33.75 | 14.50 | 29.26 | 21.69 | 0.39 | 239 |

*1(33.75) + 2(14.50) + 3(29.26) + 4(21.69) + 5(.39) = 33.75 + 29 + 87.78 + 86.76 + 1.95 = 239*

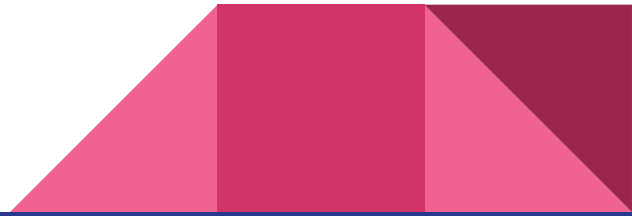
# Process (Part 1)

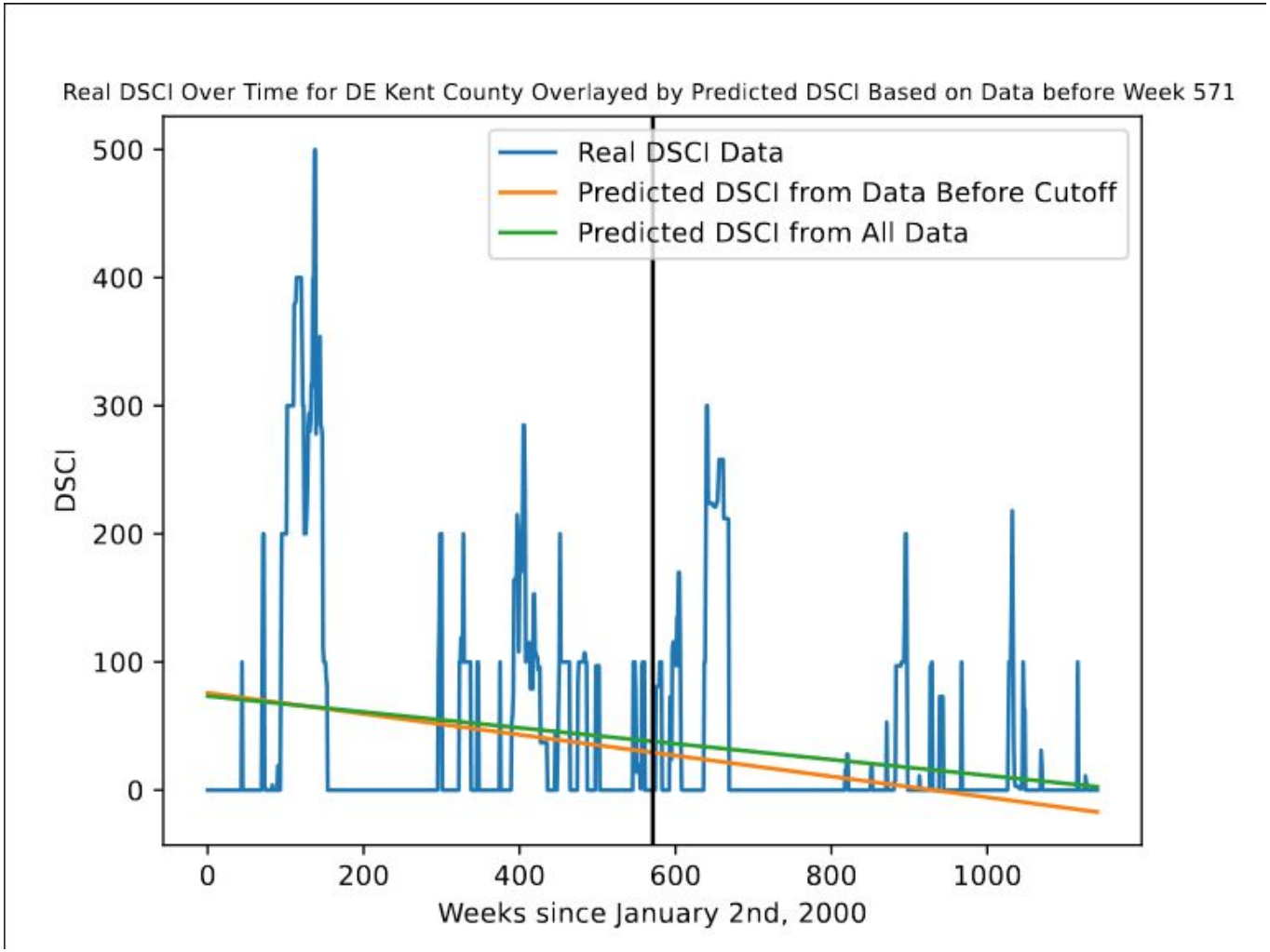
- Take the first half of the drought data, roughly from 2000 to 2010
- Use the analytic solution to find the best linear approximation of the DSCI for this portion of the dataset
- Use the analytic solution again to find the best linear approximation for the whole dataset
- Find the RSS for both models
- Visualize
- Repeat with the first three-quarters of the data

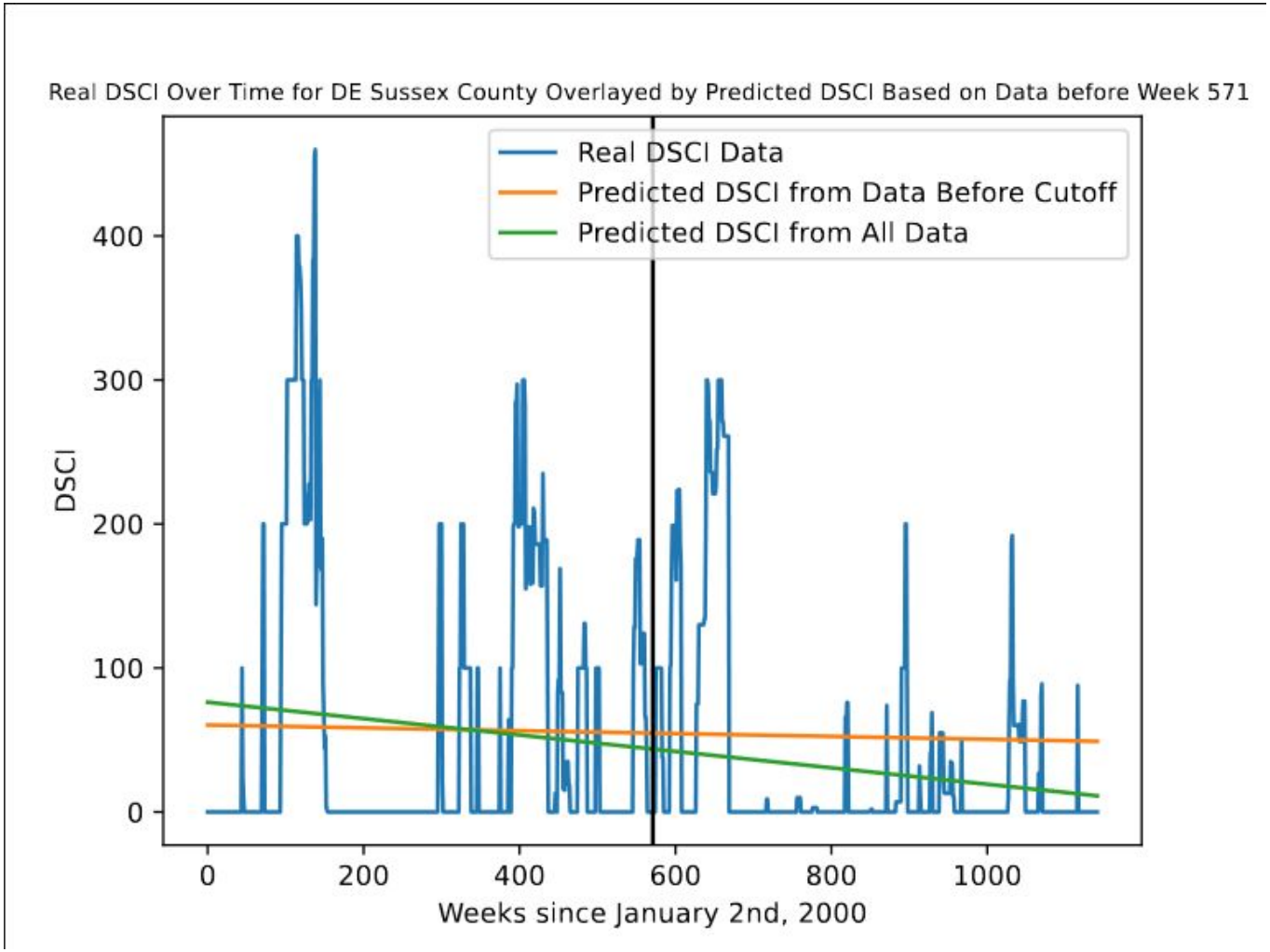


# Our Hypothesis (Part 1)

- We would be able to predict future trends of DSCI using the analytic solution on (an earlier part of) 20 years worth of DSCI indexes




Real DSCI Over Time for DE Kent County Overlayed by Predicted DSCI Based on Data before Week 571
Real DSCI Data
Predicted DSCI from Data Before Cutoff
Predicted DSCI from All Data
DSCI
500
400
300
200
100
0
0
200
400
600
800
1000
Weeks since January 2nd, 2000

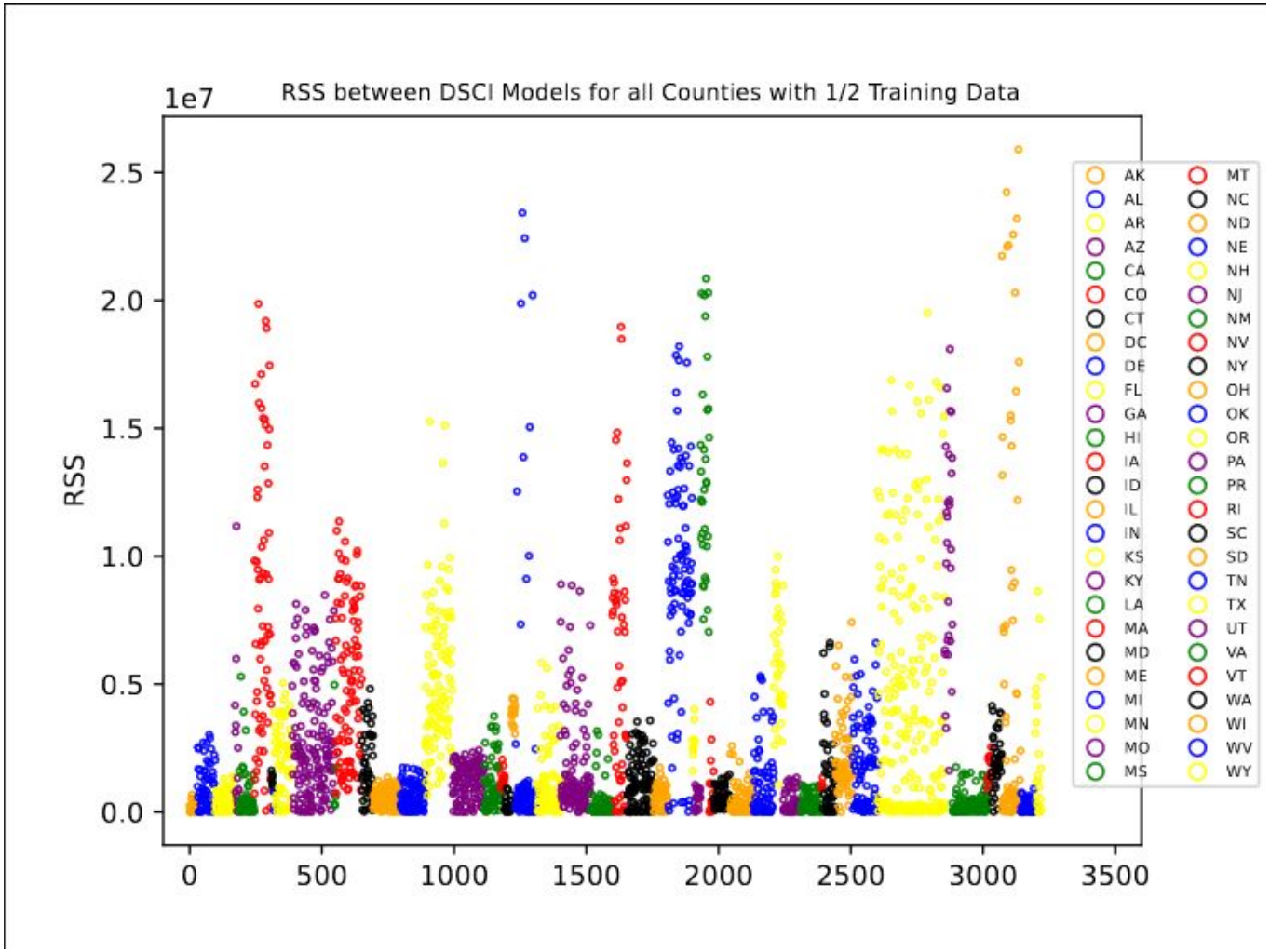

Real DSCI Over Time for DE Sussex County Overlayed by Predicted DSCI Based on Data before Week 571
Real DSCI Data
Predicted DSCI from Data Before Cutoff
Predicted DSCI from All Data
DSCI
0
100
200
300
400
0
200
400
600
800
1000
Weeks since January 2nd, 2000

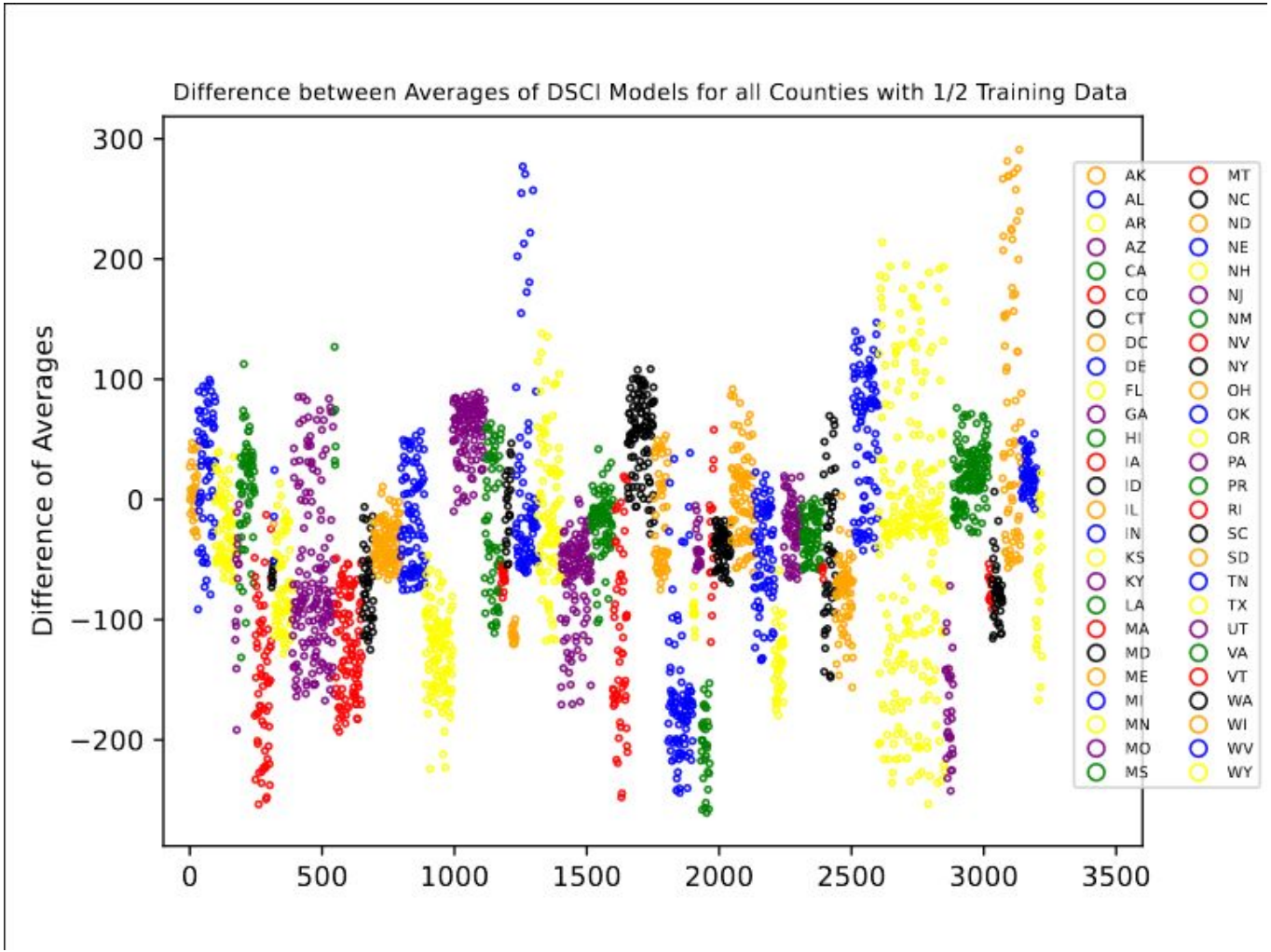
# How Accurate were our Results? (Part 1)

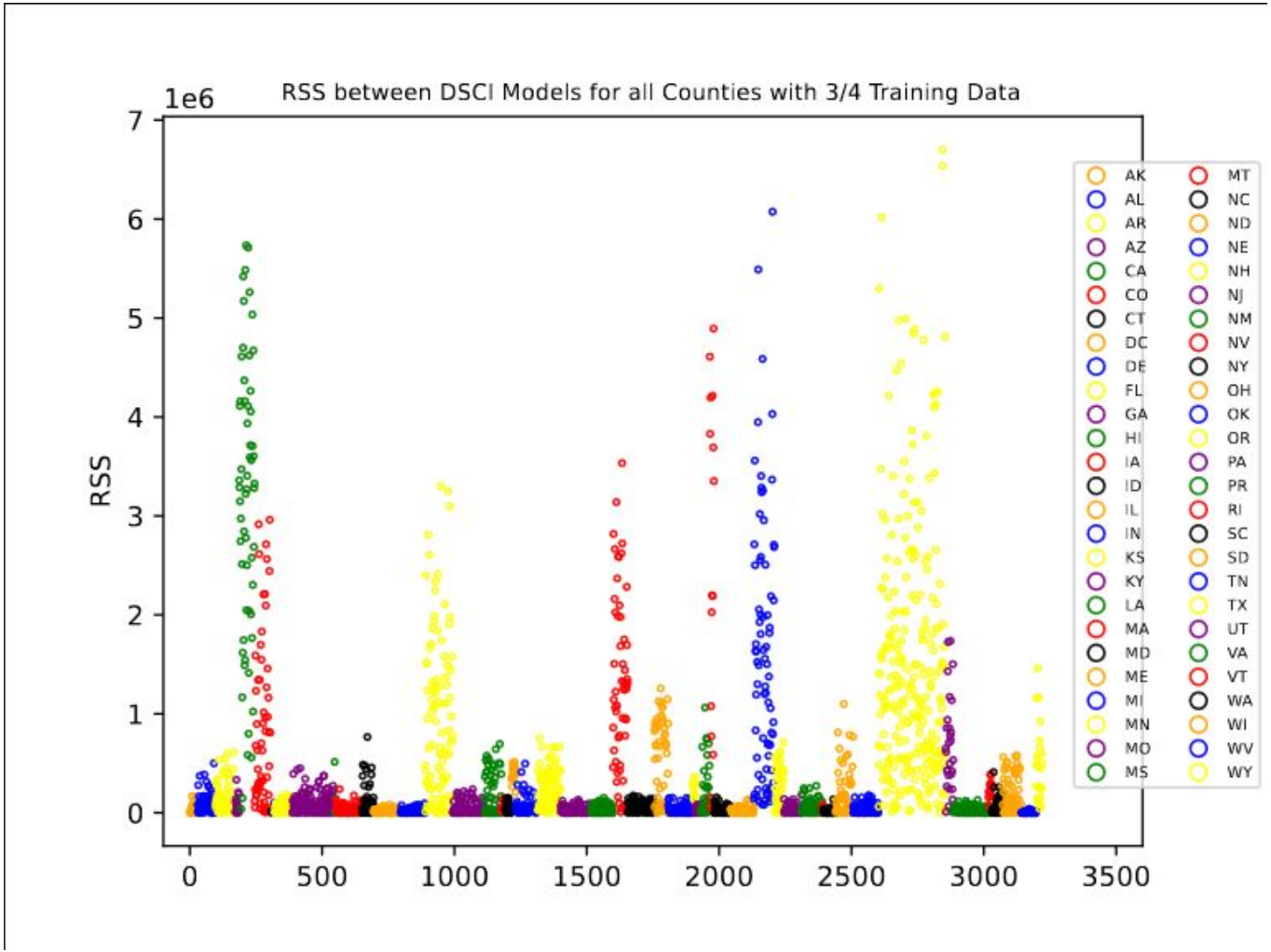
- We tried two training partitions: one with half of the data, one with three-quarters of the data

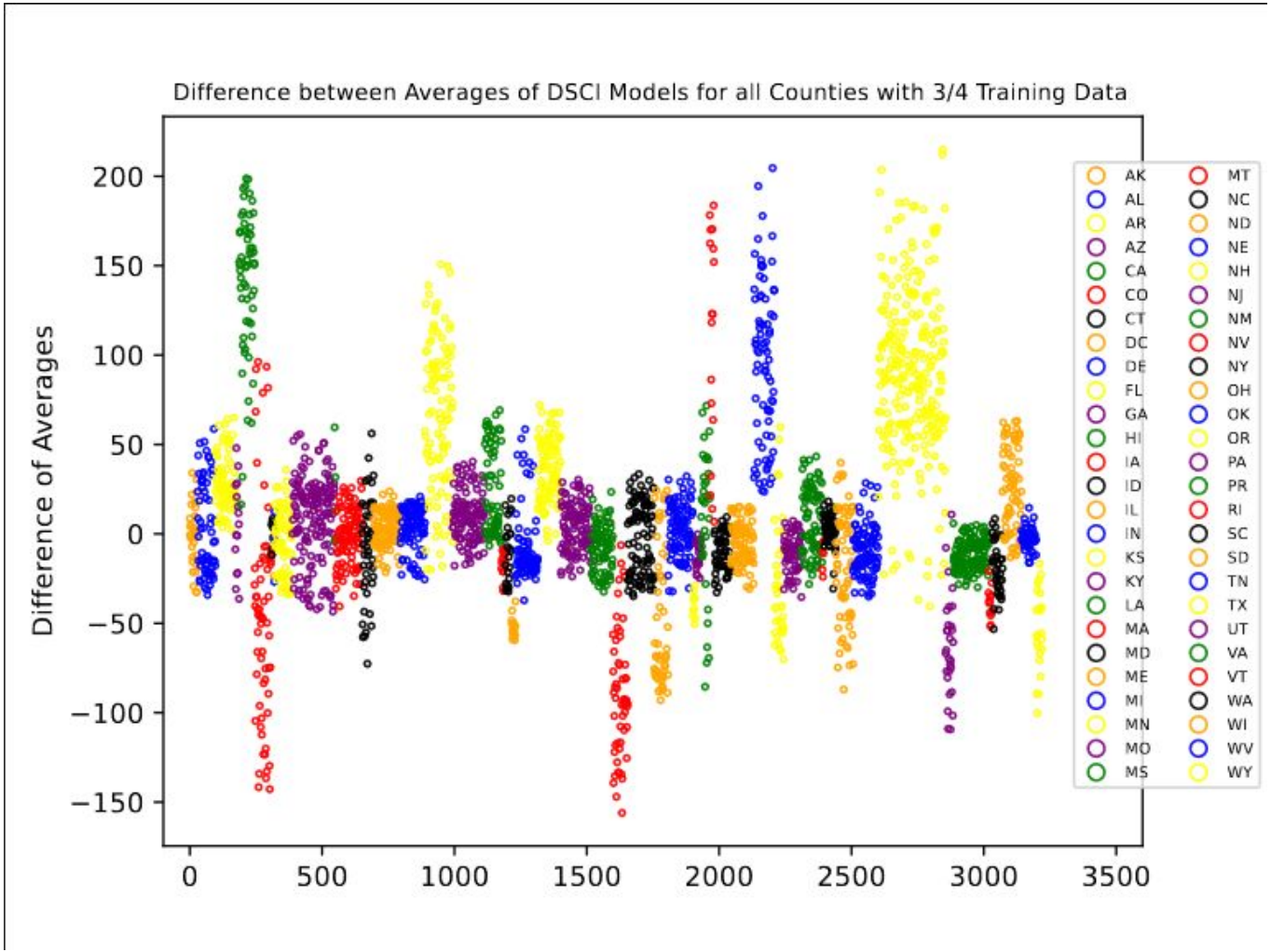
- We then calculated the RSS between a model trained on only the training data and on the entire data, for both training partitions

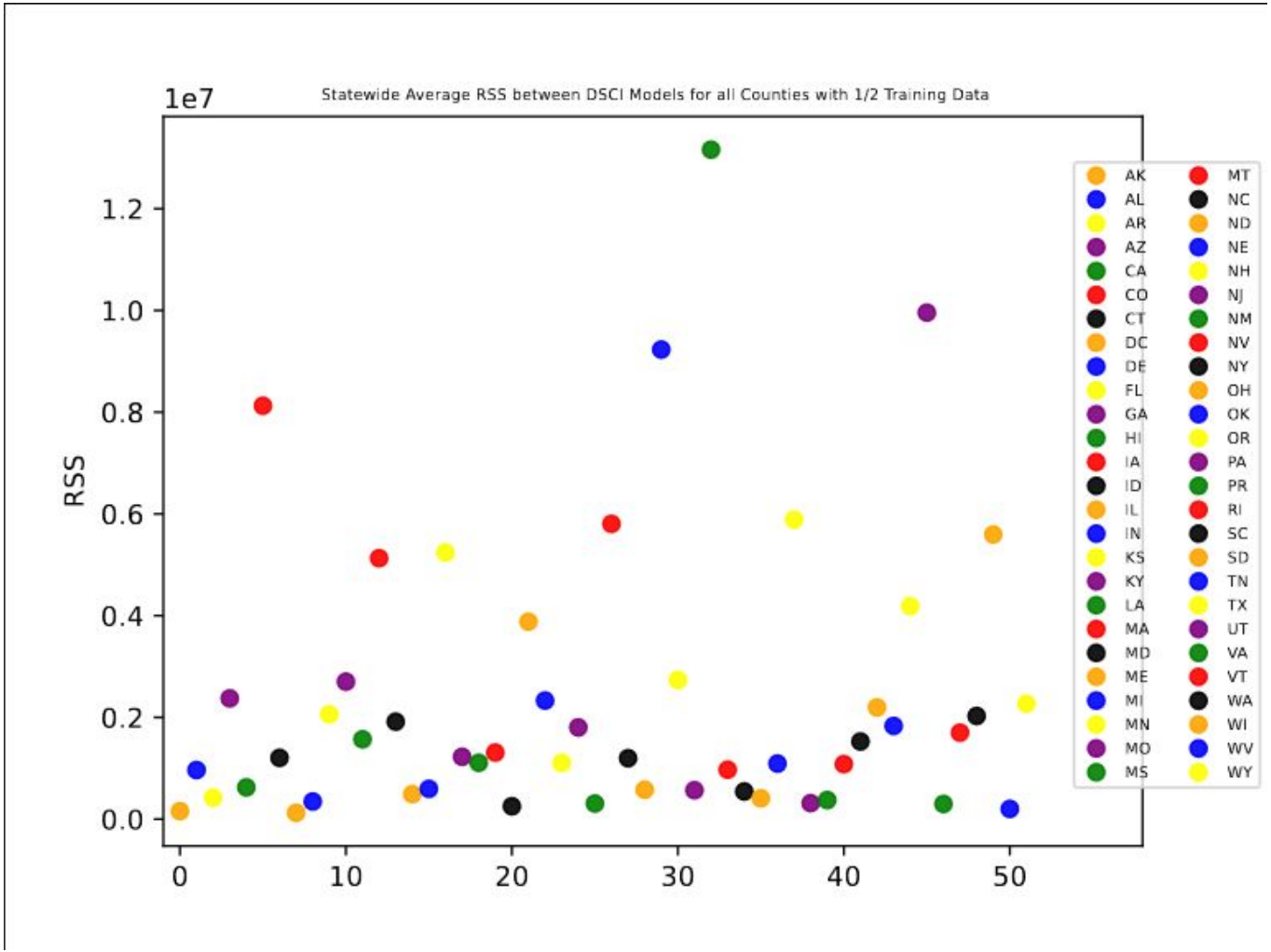
- In addition, in order to determine whether the model using partial data was underestimating or overestimating, we calculated the difference between the average DSCI for each model

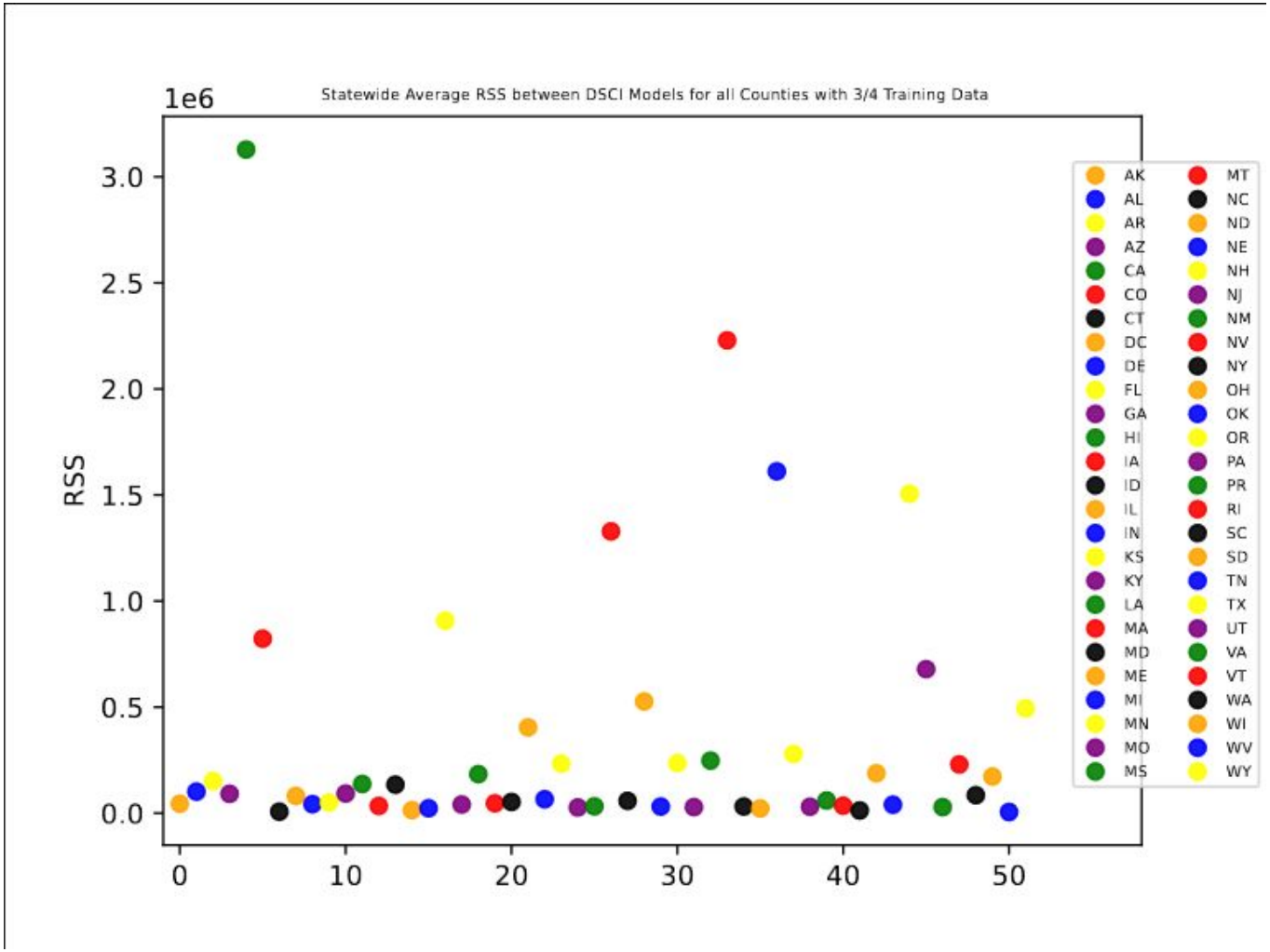

RSS between DSCI Models for all Counties with 1/2 Training Data
1e7
RSS
2.5
2.0
1.5
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
3500
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY

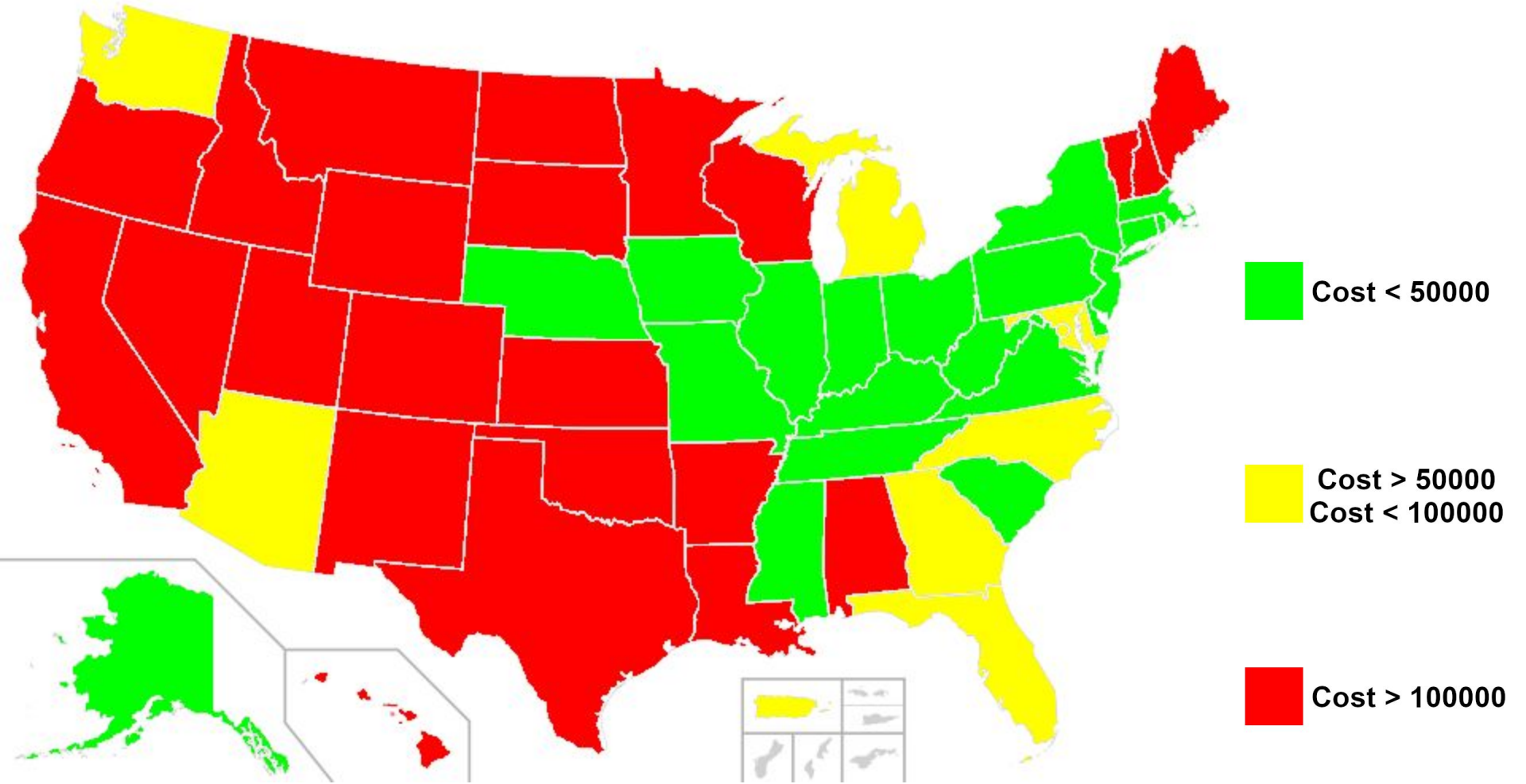

Difference between Averages of DSCI Models for all Counties with 1/2 Training Data
Difference of Averages
300
200
100
0
−100
−200
0
500
1000
1500
2000
2500
3000
3500
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY


RSS between DSCI Models for all Counties with 3/4 Training Data
1e6
RSS
0
1
2
3
4
5
6
7
0
500
1000
1500
2000
2500
3000
3500
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY


Difference between Averages of DSCI Models for all Counties with 3/4 Training Data
Difference of Averages
200
150
100
50
0
−50
−100
−150
0
500
1000
1500
2000
2500
3000
3500
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY

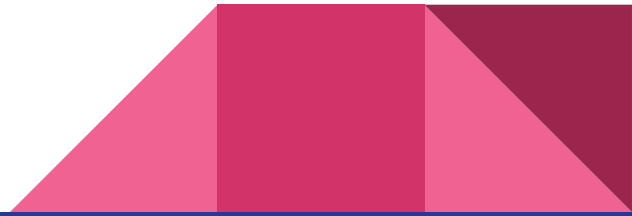

Statewide Average RSS between DSCI Models for all Counties with 1/2 Training Data
1e7
RSS
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
10
20
30
40
50
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY


Statewide Average RSS between DSCI Models for all Counties with 3/4 Training Data
1e6
RSS
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
PR
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY


Cost < 50000
Cost > 50000
Cost < 100000
Cost > 100000

# Our Hypothesis (Part 2)

- Naive Bayes will show us that there is a correlation between median household income and DSCI indexes



# Does a county's drought affect it's median income?

- Used Naive Bayes to see correlation between median household income and DSCI of a county
- Low income: Median household income ≤ $51,852
- Low DSCI: DSCI ≤ average DSCI of counties being viewed
- Shuffled the data and divided it in half to get a training and testing partition
- Repeated 20 times and took the average accuracy to get a more robust result
- Accuracy of ~.55 which is a little more that coin toss
- Naive Bayes shows DSCI indexes have a very small effect on the median household income of a county

# Conclusions

- For the most part, the rate at which the amount of drought in a state increases or decreases is very nonlinear (i.e we can't predict it using only previous data)
- However, how erratic the rate is varies substantially by county and geographical region
- There is only a weak correlation between a county's median income and how much drought it's in

# Expansions

- Gather median household data from all counties over a wider span of years
- Create a map with every county categorized, instead of every state