

It Takes a Village

ELI WINTON, EMILY K. SANDERS, RADHA MOHANTY

*A Presentation for the Coalition for Unprecedented Transformative
Education for California Teachers and Students, 2024 Meeting*

Problem Statement

- ▶ Dropping out of high school impairs a child's opportunities for the rest of their life.
- ▶ You all want to help!
- ▶ We used government data to predict dropout rates from social conditions at the county level in California.
- ▶ Goals: Plan community actions, set political goals

Definitions

Graduation Cohort

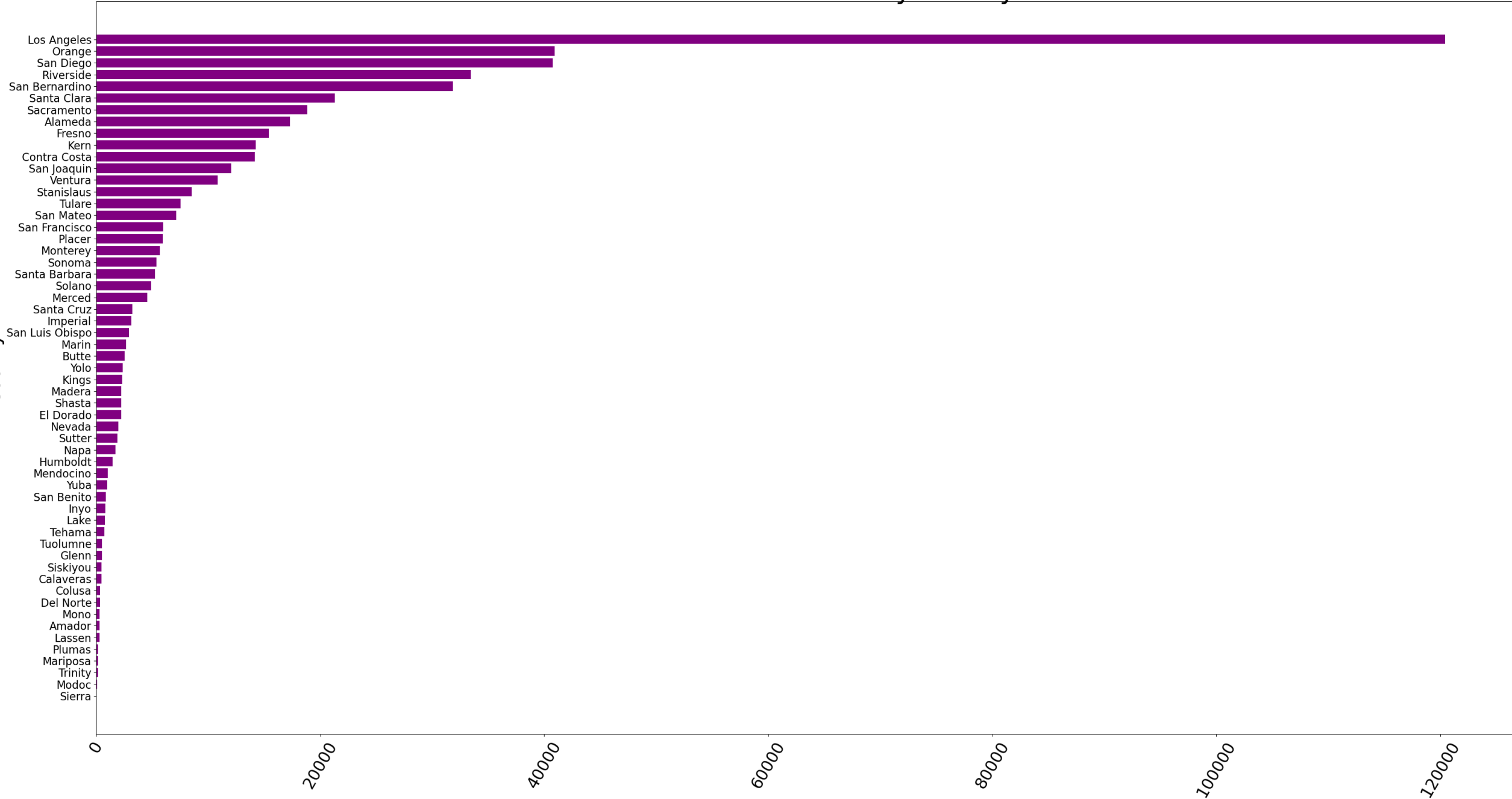
- ▶ All students who enter 9th grade at the same time
- ▶ Plus/minus students who leave or join the school
- ▶ Public school only

Dropout Rate

- ▶ Percentage of the graduation cohort who:
- ▶ Do not complete in any way
- ▶ Are not still enrolled

2019 Graduation Cohort Size by County

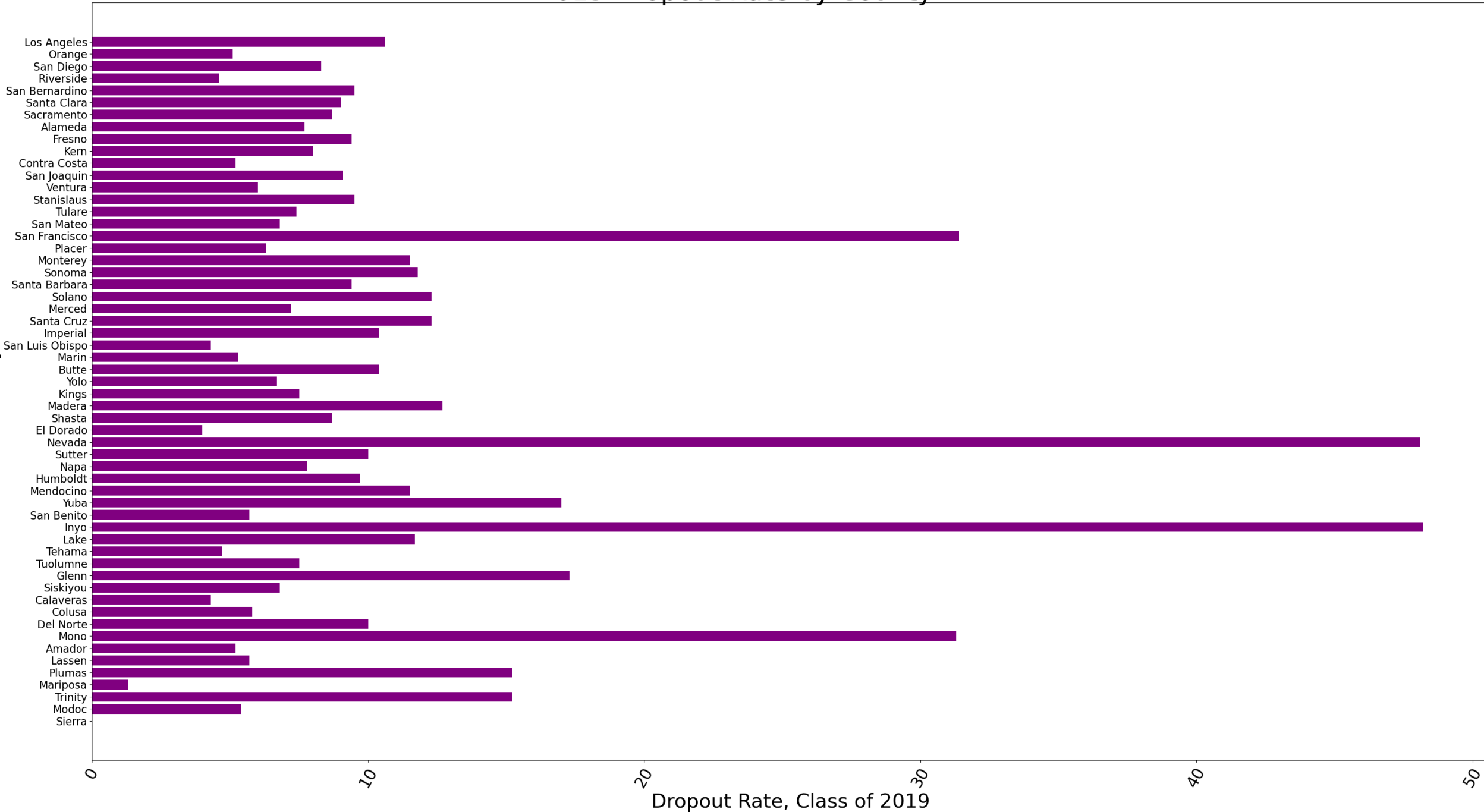
County



Graduation Cohort Size, Class of 2019

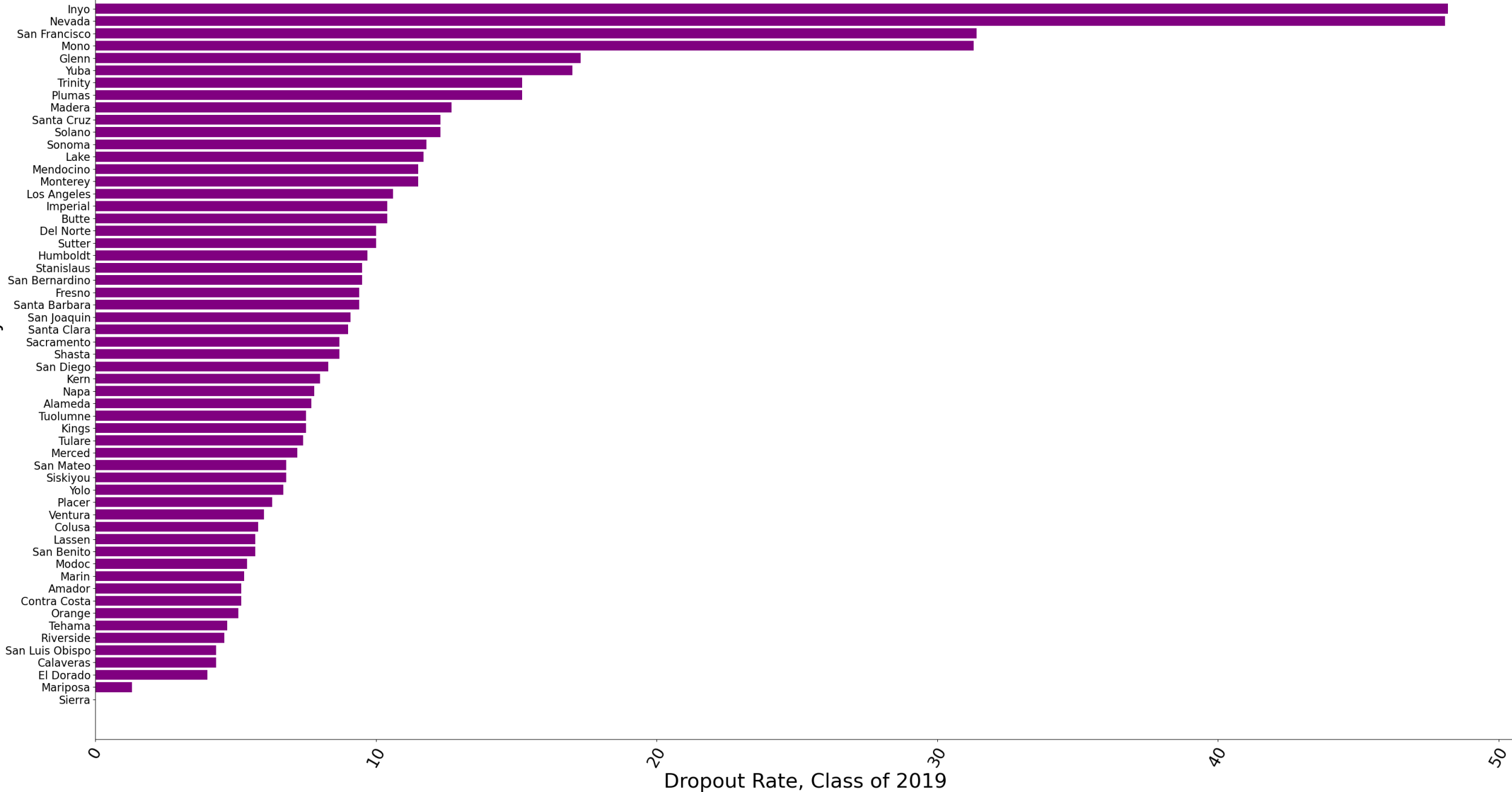
2019 Dropout Rate by County

County



2019 Dropout Rate by County

County

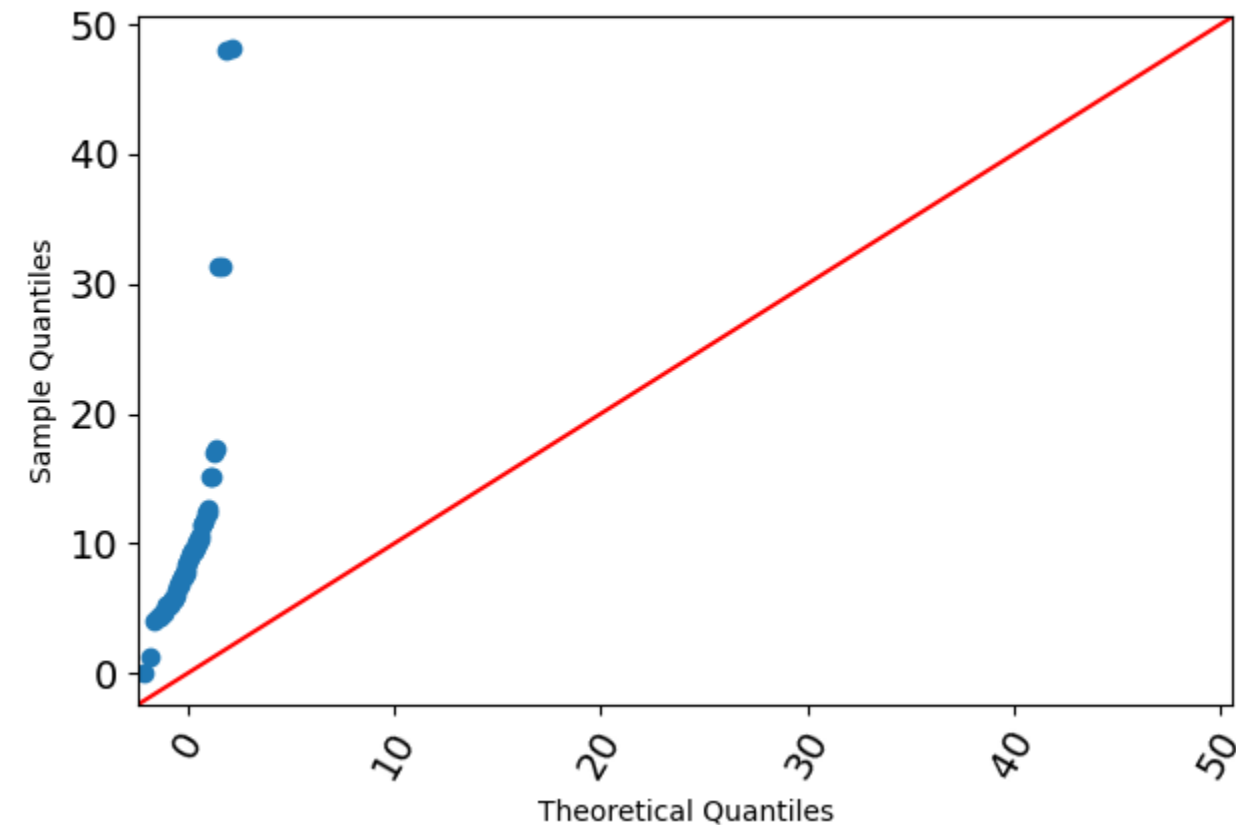


Method

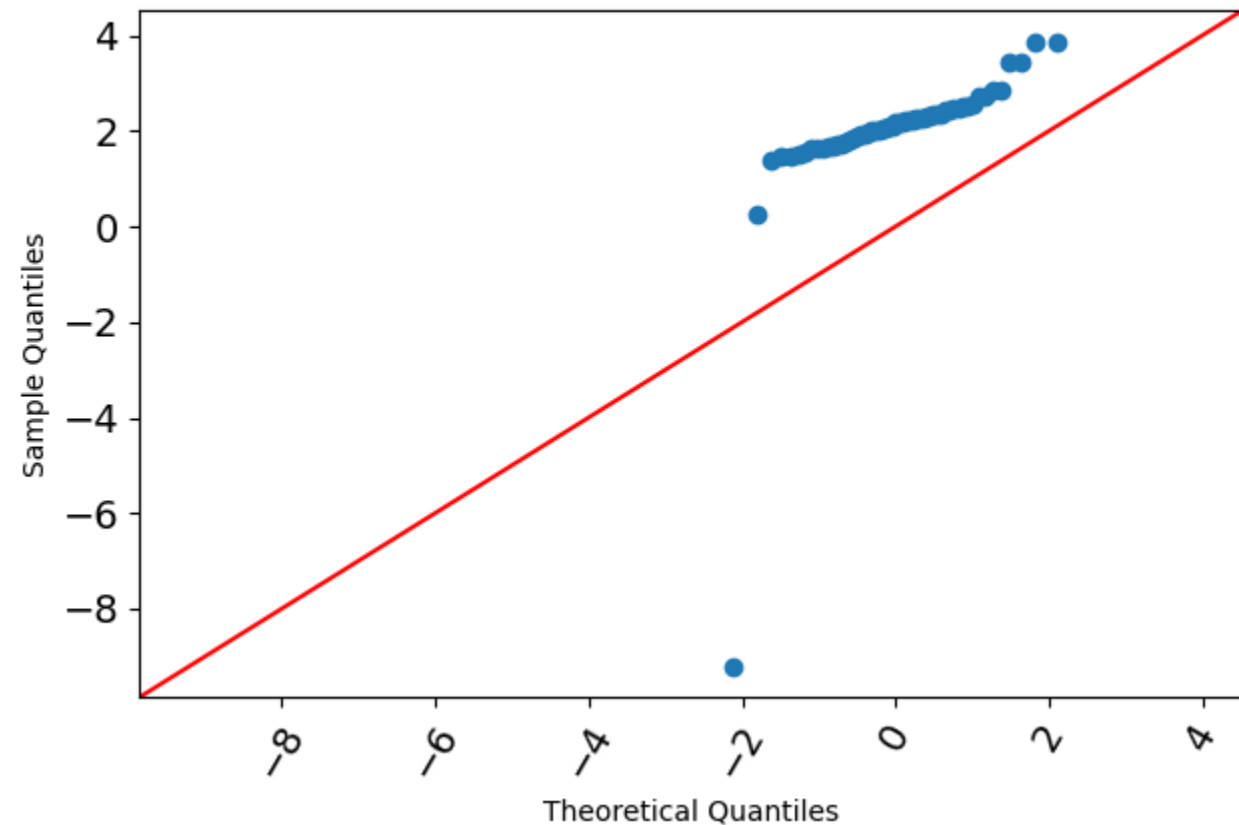
- ▶ 14 government datasets
- ▶ 1 row per county, 100s of features (at first)
- ▶ All predictors came from ~2015, dropout rate from 2019
- ▶ Feature engineering: condense, impute, polynomialize
- ▶ Feature selection: correlations, LASSO

Distribution Woes

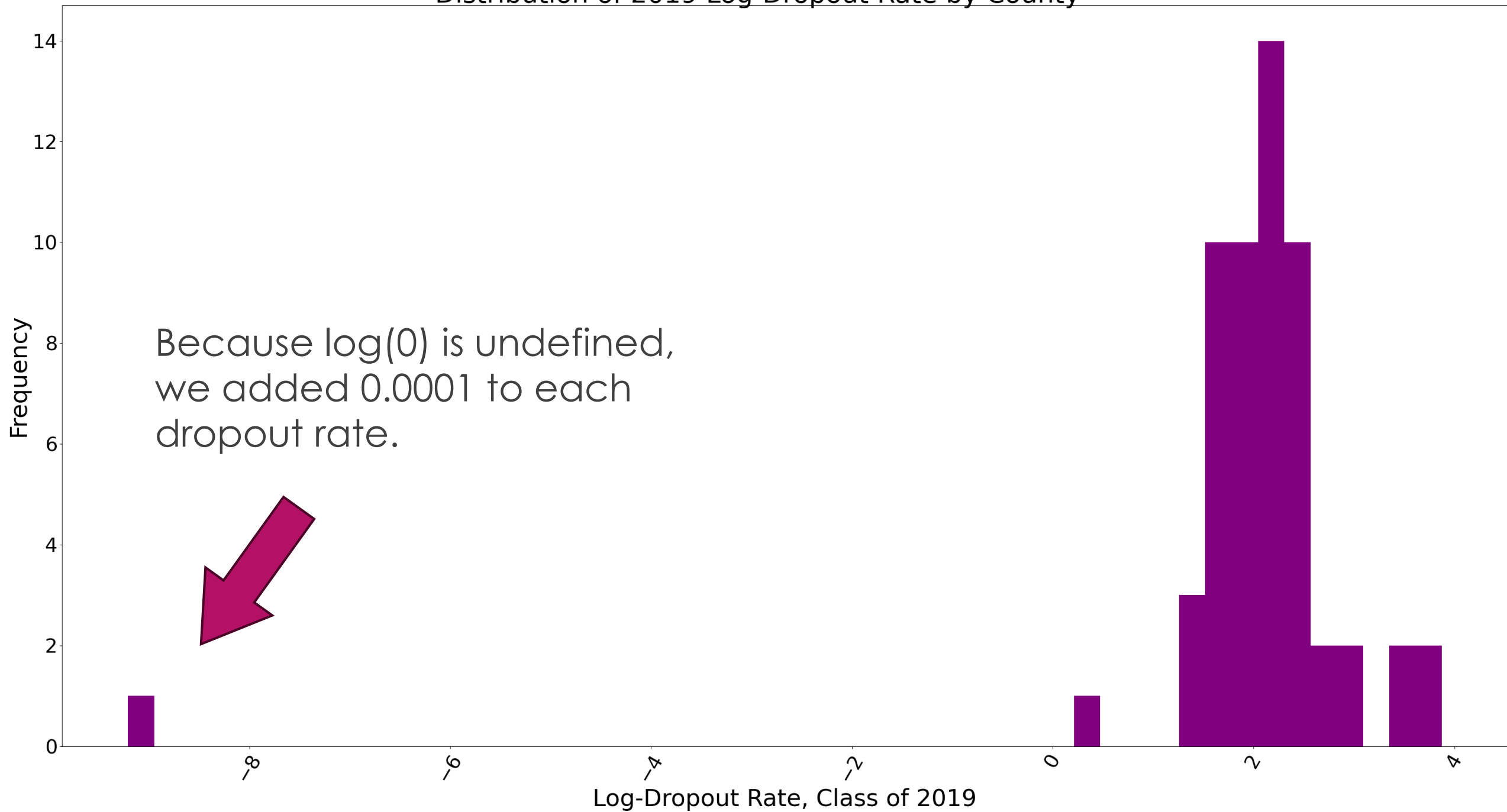
QQ-Plot of 2019 Dropout Rate by County



QQ-Plot of 2019 Log Dropout Rate by County



Distribution of 2019 Log-Dropout Rate by County

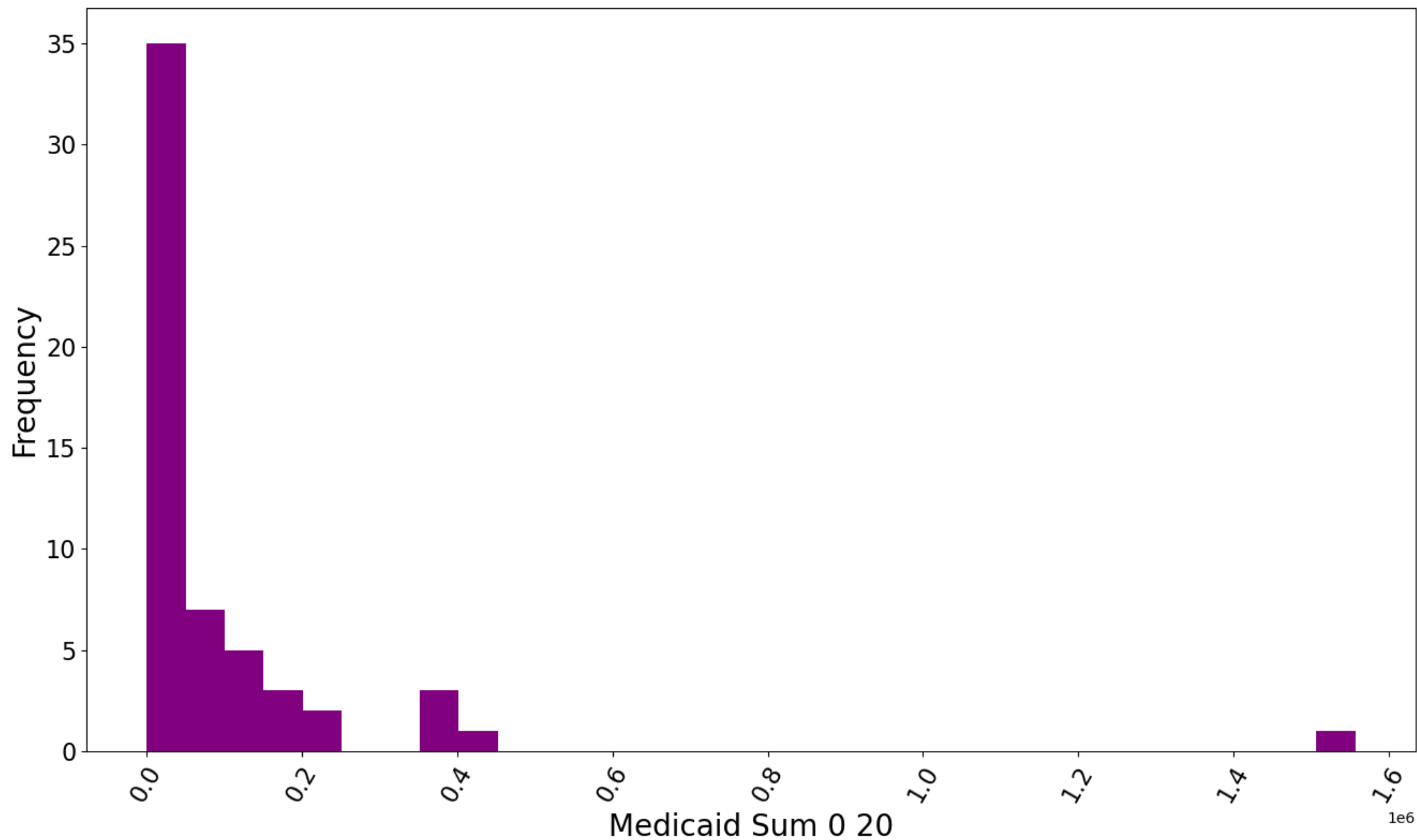


Features

- ▶ Population & cohort size
- ▶ Poverty & income inequality
- ▶ Unemployment rates
- ▶ Medicaid enrollees <21yo
- ▶ Racial demographics <6yo
- ▶ Teenage birth rate
- ▶ Availability of ecigarettes
- ▶ STD rates
- ▶ Suicide, asthma death rates
- ▶ Infant mortality
- ▶ Medicaid abortion funding
- ▶ Daycare availability

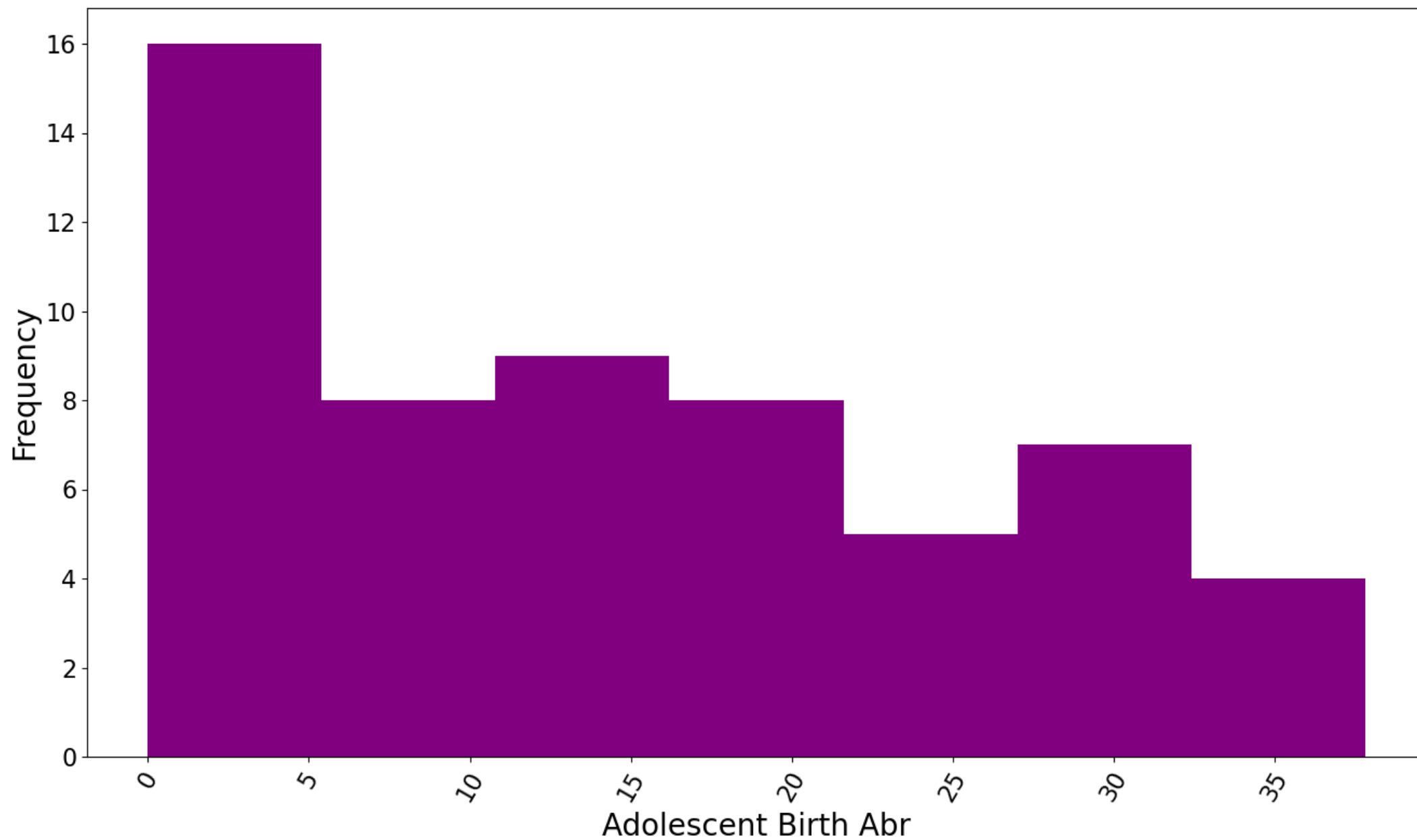
Distribution of Medicaid Sum 0 20

Based on 57 Observations out of 57



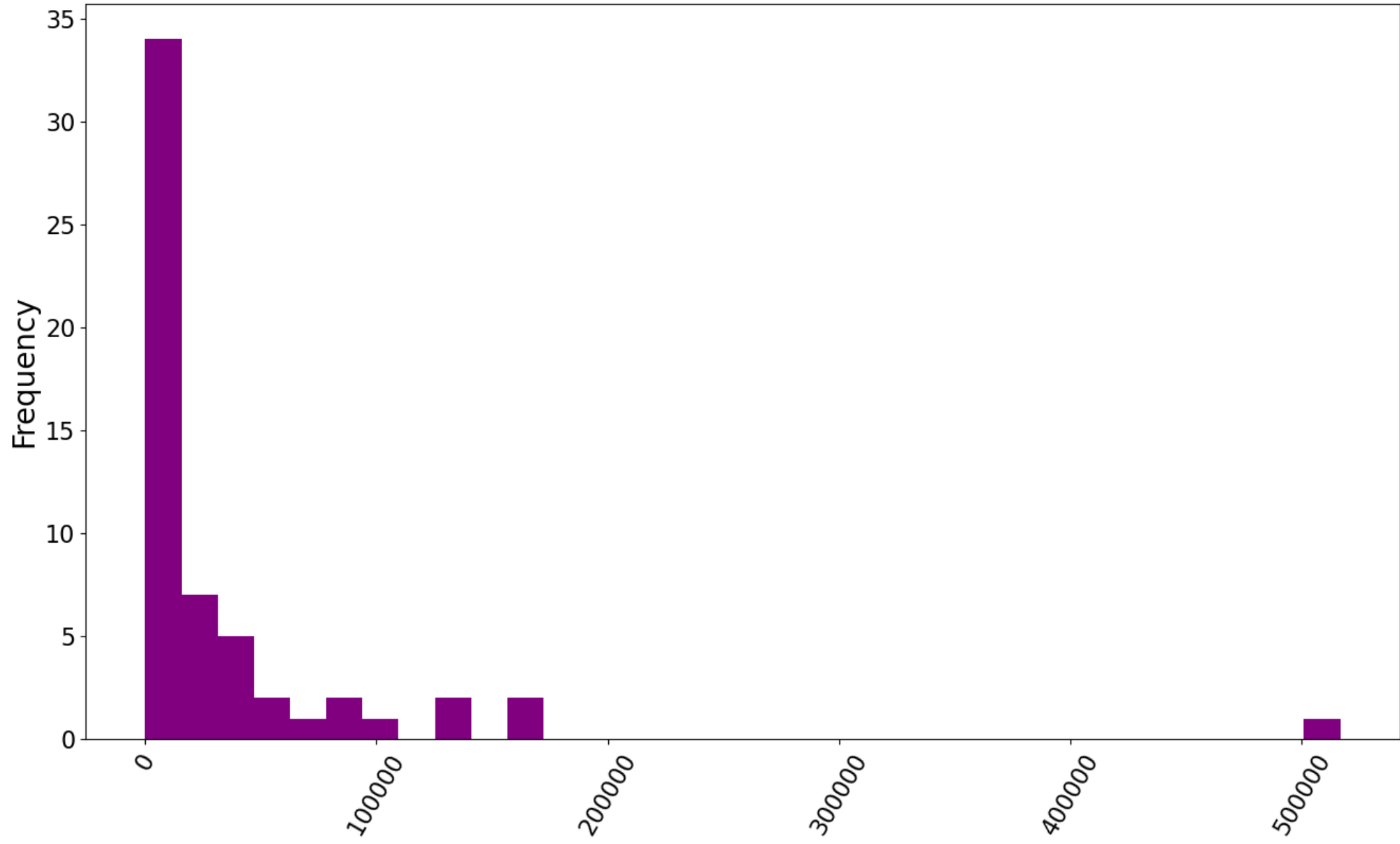
Distribution of Adolescent Birth Abr

Based on 57 Observations out of 57



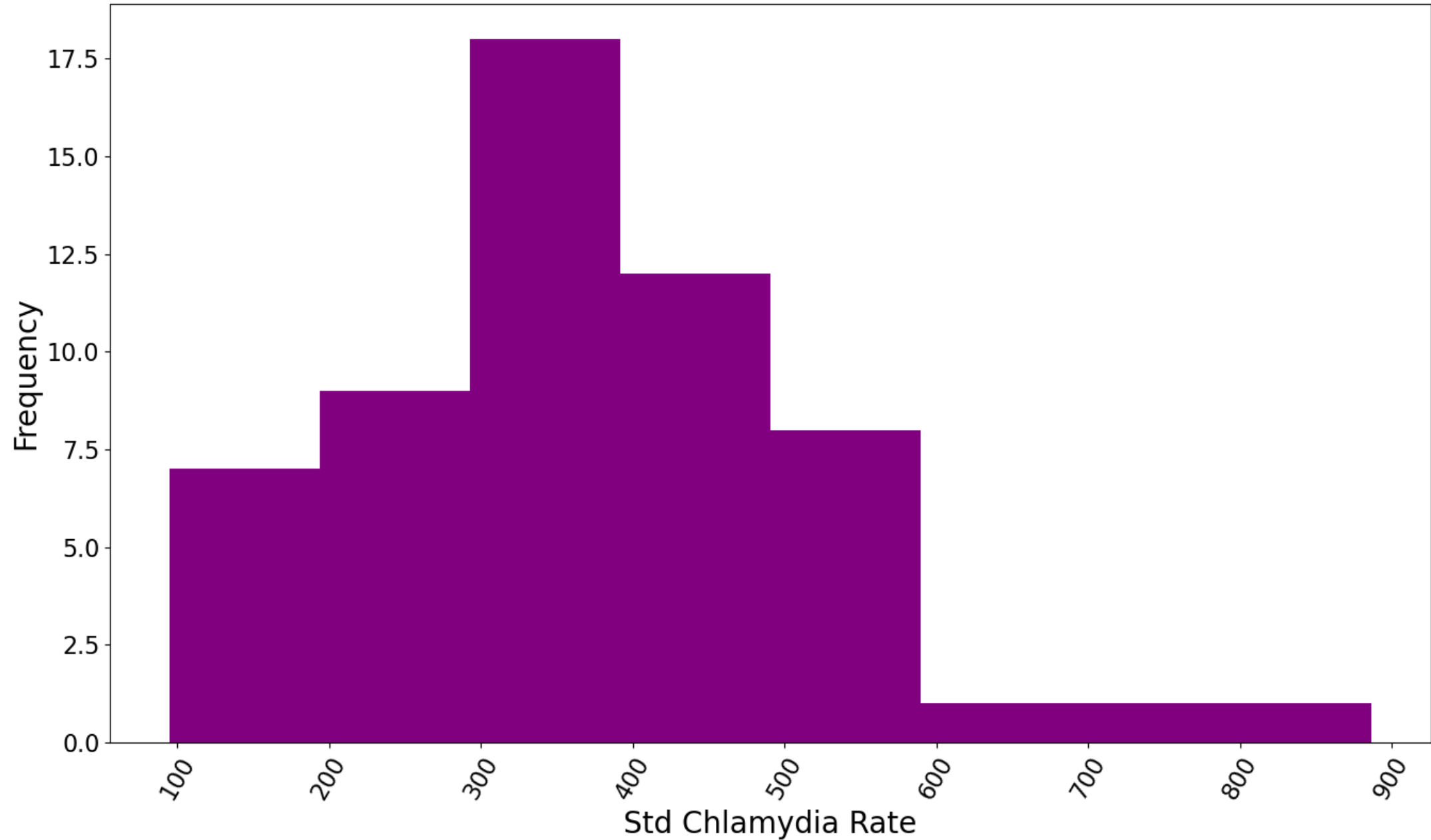
Distribution of Daycare Slots Child Total Pop

Based on 57 Observations out of 57



Distribution of Std Chlamydia Rate

Based on 57 Observations out of 57



Results

- ▶ Very challenging to find a good model
- ▶ The best we found was

▶ **WRITE ABOUT AT
LEAST ONE MODEL**

Results

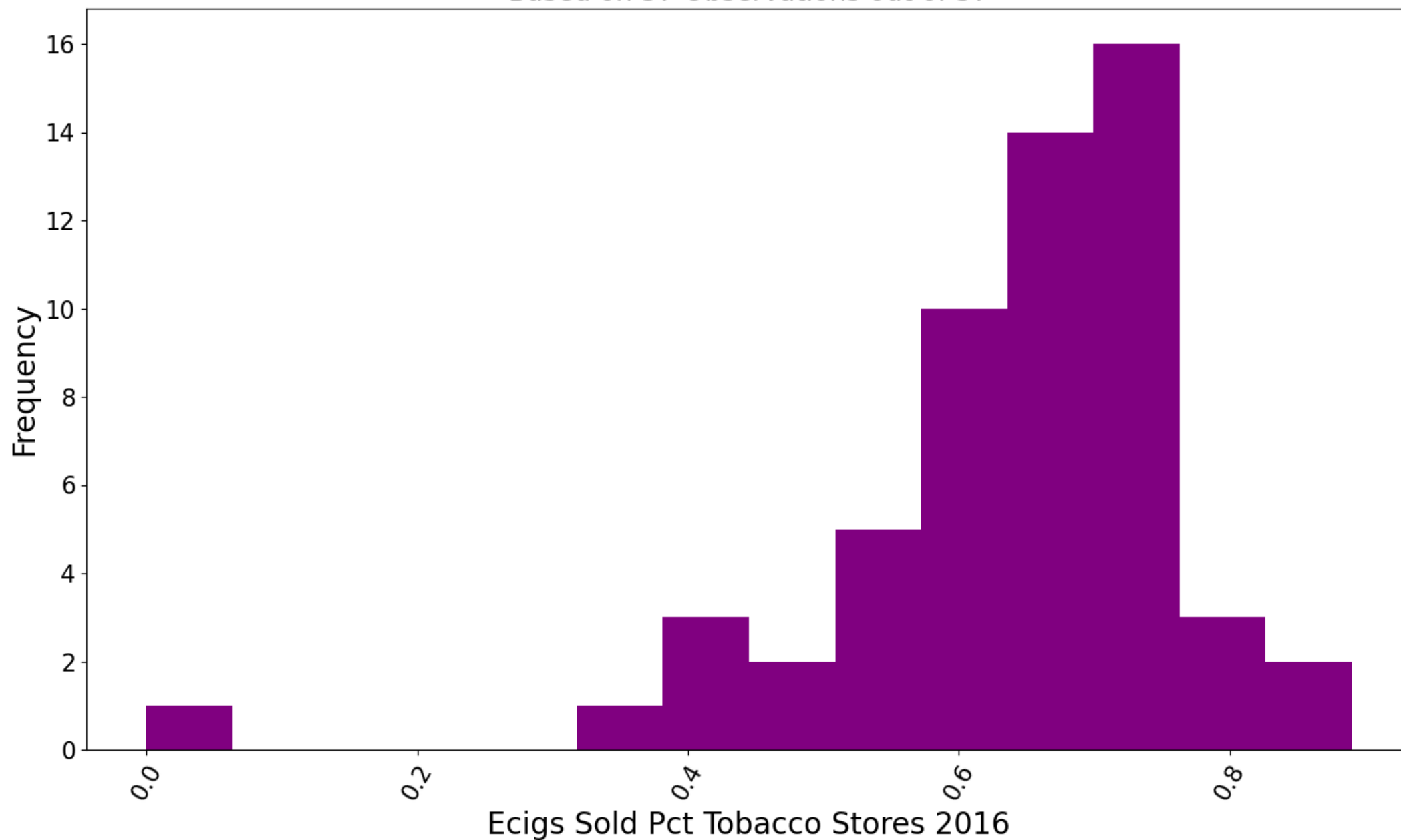
- ▶ Correlations do not imply causation, but...
- ▶ Availability of ecigarettes seems bad
- ▶ Moreso in combination with other factors than alone



Photo credit: [Vaping360](#), [CC BY 2.0](#)

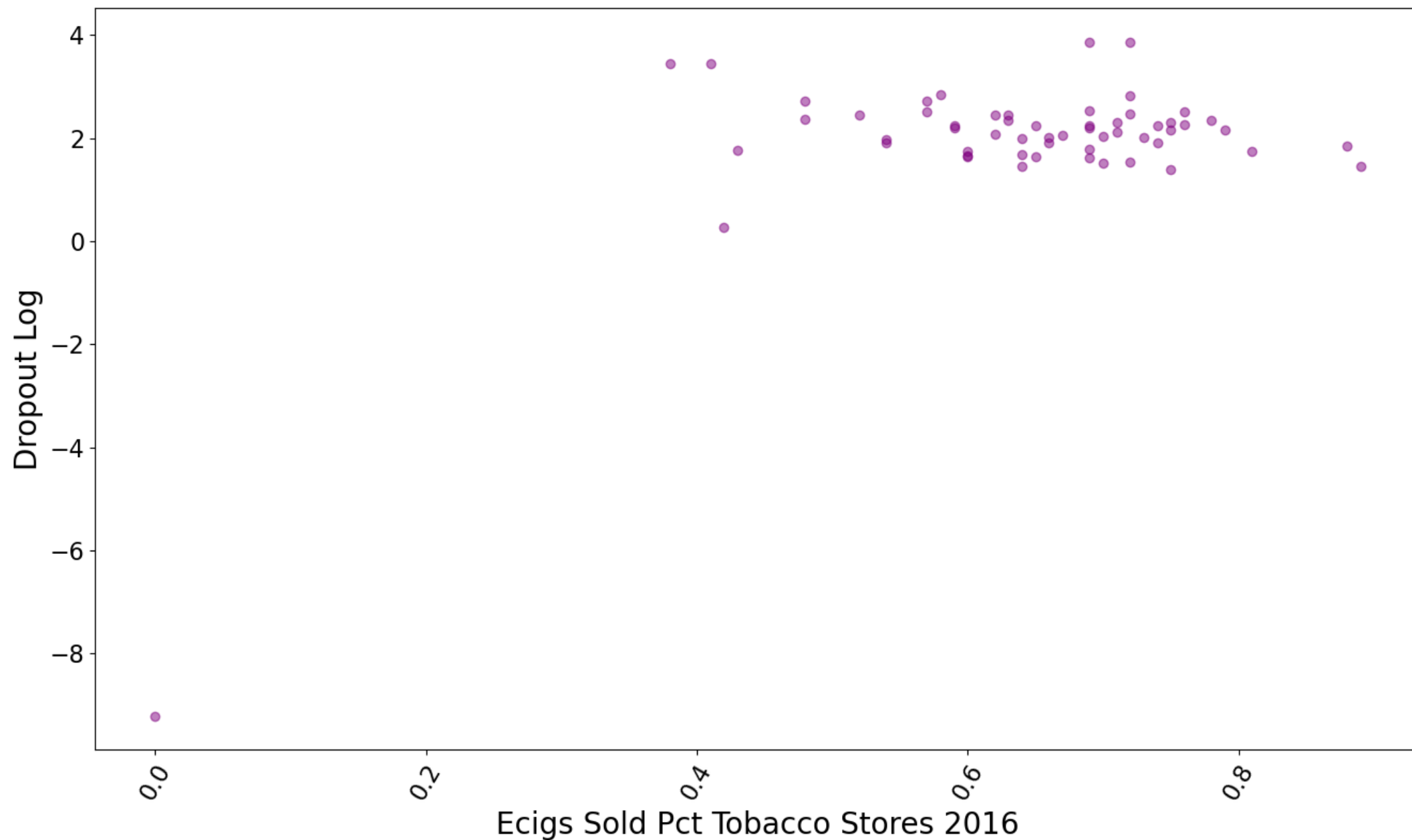
Distribution of Ecigs Sold Pct Tobacco Stores 2016

Based on 57 Observations out of 57



Relationship between Ecigs Sold Pct Tobacco Stores 2016 and Dropout Log

Based on 57 Observations out of 57



Limitations

- ▶ Irregular data leads to faulty models
- ▶ Have to choose: exclude counties, or limit performance
- ▶ “Obvious” predictors not readily available (e.g., race)
- ▶ Many predictors available as counts, not rates
- ▶ Data is old and pre-pandemic

Recommendations

For us

- ▶ Control for population better
- ▶ Evaluate outliers separately
- ▶ Refine, drop, add features
- ▶ Different types of models
- ▶ Fresher data (Covid!)

For you

- ▶ The problem is real, so keep using your expertise!
- ▶ Ponder combined effects
- ▶ Regulate ecigarette sales

Conclusion

- ▶ The goal was to understand and reduce dropout rates.
- ▶ Our first outing was not a huge success.
- ▶ Try to reduce the availability of ecigarettes.
 - ▶ Legal regulation
 - ▶ Cultural popularity
- ▶ It'll take more work, but we're in if you're in.



Thank you!

We will now take questions.