

# How Decisions Are Made For United States Schools

## Data Sources

**BAC 2017 Dataset:**

1. federalunds.txt
2. federalundstxtheadr.xlsx
3. federalundskxy.xls
4. federalundsfundagencykey.xlsx
5. samplesocioecodata.csv

**Data We Added:**

- United States
- Zip Codes

## Federal Programs from the Agencies



## Predictive Machine Learning Algorithms

Turn Amount from Continuous to Discrete by Grouping by Intervals

```

update ff_44_May_6 set AMNTLVL = 'LE_4M'
where AMOUNT > 3000000 and AMOUNT <= 4000000;
update ff_44_May_6 set AMNTLVL = 'LE_5M'
where AMOUNT > 4000000 and AMOUNT <= 5000000;
update ff_44_May_6 set AMNTLVL = 'LE_6M'
where AMOUNT > 5000000 and AMOUNT <= 6000000;
update ff_44_May_6 set AMNTLVL = 'LE_7M'
where AMOUNT > 6000000 and AMOUNT <= 7000000;
update ff_44_May_6 set AMNTLVL = 'LE_8M'
where AMOUNT > 7000000 and AMOUNT <= 8000000;
update ff_44_May_6 set AMNTLVL = 'LE_9M'
where AMOUNT > 8000000 and AMOUNT <= 9000000;
update ff_44_May_6 set AMNTLVL = 'LE_10M'
where AMOUNT > 9000000 and AMOUNT <= 10000000;
update ff_44_May_6 set AMNTLVL = 'LE_10M'
where AMOUNT > 10000000 and AMOUNT <= 150000000;
update ff_44_May_6 set AMNTLVL = 'L25M_100M'
where AMOUNT > 25000000 and AMOUNT <= 100000000;
update ff_44_May_6 set AMNTLVL = 'L25M_100M'
where AMOUNT > 250000000 and AMOUNT <= 1000000000;
update ff_44_May_6 set AMNTLVL = 'L100M_1B'
where AMOUNT > 100000000 and AMOUNT <= 1000000000;
update ff_44_May_6 set AMNTLVL = 'L100M_1B'
where AMOUNT > 1000000000 and AMOUNT <= 10000000000;
update ff_44_May_6 set AMNTLVL = 'LE_9M'
where AMOUNT > 8000000 and AMOUNT <= 9000000;
update ff_44_May_6 set AMNTLVL = 'L1B_Max'
where AMOUNT > 1000000000;
    
```

## Decision Tree Models

**Predicting Amount when**

```

ffs <- subset(ff, ff$PER_0_10 > 4.652)
ff$AMOUNT <- ff$AMOUNT/100000
ff$AMOUNT <- ctree(data=ffs, AMOUNT~PER_WHITE+PER_POVERTY+
PER_HS_GRAD+PER_BACH_DEGREE+
PER_0_10+PER_10_19)
    
```

**Predicting Amount when**

```

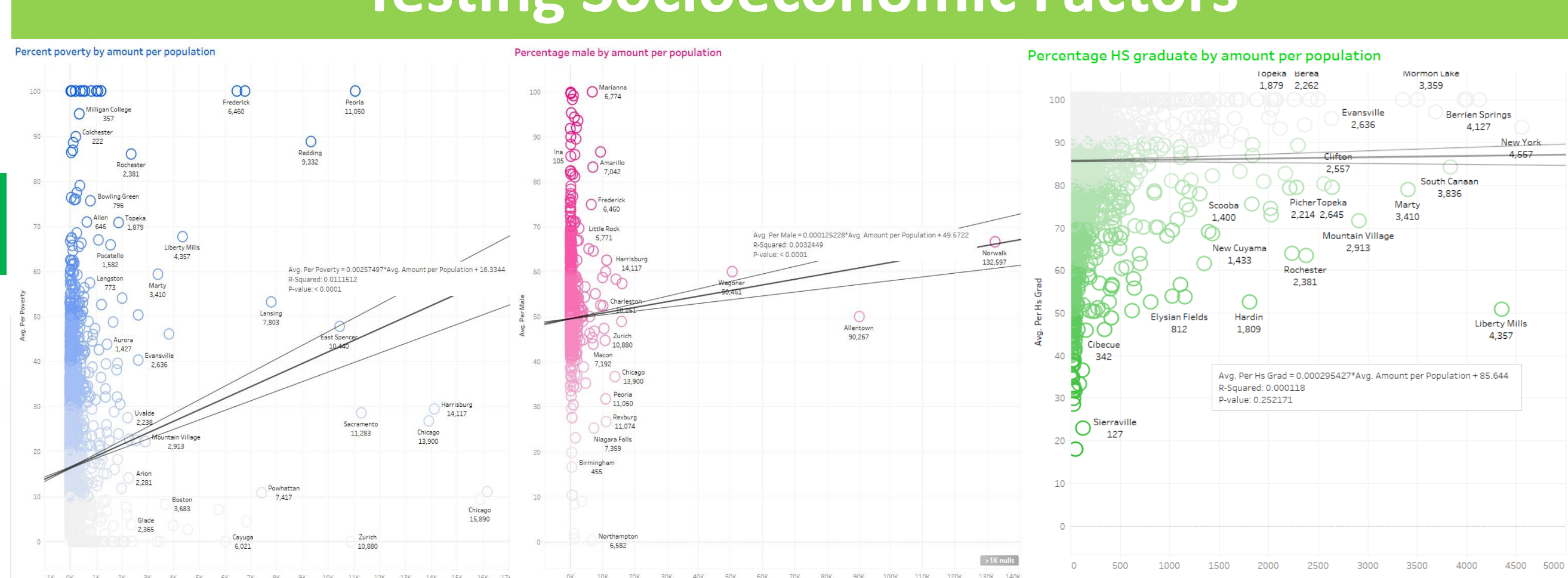
ffss <- subset(ff, ff$PER_0_10 > 4.652)
& (ff$PER_0_10 > 9.693) & (ff$PER_POVERTY <= 29.14)
& (ff$PER_WHITE <= 75.762) & (ff$PER_10_19 > 5.059)
& (ff$PER_POVERTY <= 75) & (ff$PER_WHITE < 59)
& (ff$PER_POVERTY > 18.487) & (ff$PER_10_19 <= 20.143)
ff$AMOUNT <- ctree(data=ffss, AMOUNT~PER_WHITE+PER_POVERTY+
PER_HS_GRAD+PER_BACH_DEGREE+
PER_0_10+PER_10_19)
    
```

## Cleaning the Data in R and SQL

```

select distinct FEDERALPROGRAMNAME from [dbo].[FF_42_Unk]
where FEDERALPROGRAMNAME <> ''
from [dbo].[FF_42_Unk] where (CATEGORIC) <> 'BKGST001' -- 6993
select count(*) from [dbo].[FF_42_Unk]
where FEDERALPROGRAMNAME <> ''
from [dbo].[FF_42_Unk] where (CATEGORIC) <> 'BKGST001' -- 32394
select (CATEGORIC), count(*) from [dbo].[FF_42_Unk]
group by (CATEGORIC)
select count(*) from FF_42
select * into FF_42_Cat from FF_42 where CATEGORIC = 'unk'
select count(*) from FF_42_Cat
select count(*) from FF_42_Cat
select distinct CATEGORIC from FF_42
where agency = 'Department of Agriculture'
select FEDERALPROGRAMNAME, count(*) as count
from FF_42
where CATEGORIC = 'B' and amount < 10000
group by FEDERALPROGRAMNAME order by count desc)
select FEDERALPROGRAMNAME, count(*) as count
from FF_42
where CATEGORIC = 'B' and amount < 10000
group by FEDERALPROGRAMNAME order by count desc)
select FEDERALPROGRAMNAME, count(*) as count
from FF_42
where CATEGORIC = 'B' and amount < 10000
group by FEDERALPROGRAMNAME order by count desc)
    
```

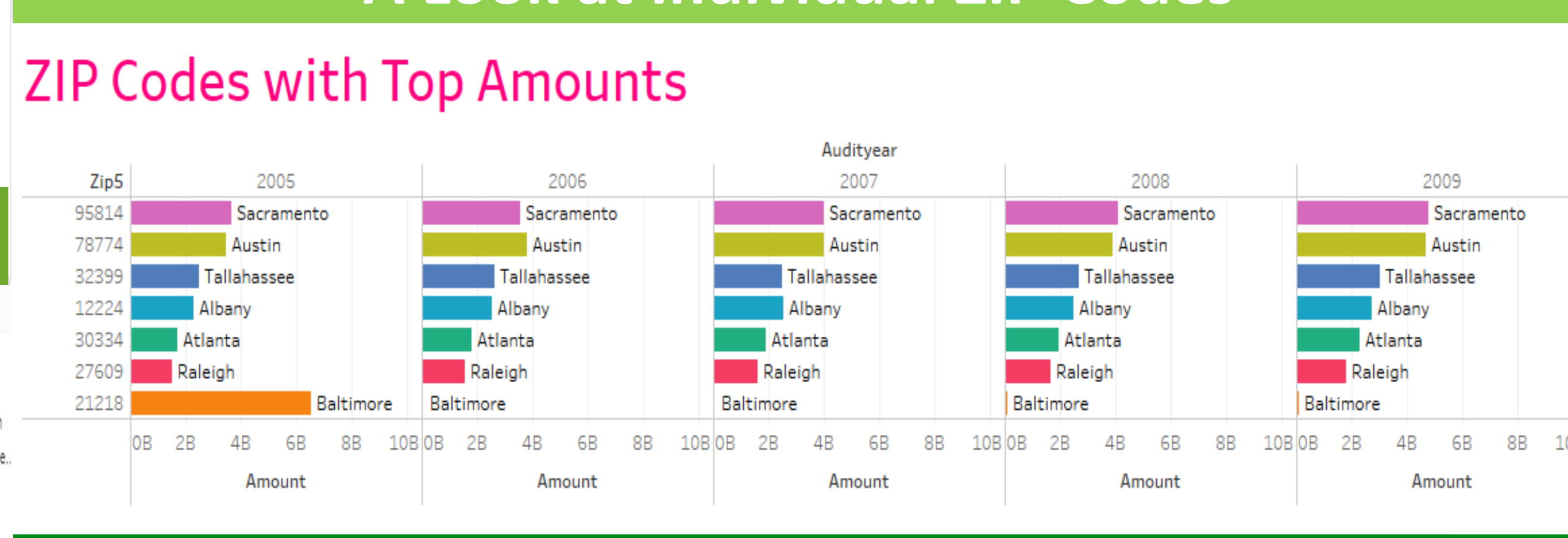
## Testing Socioeconomic Factors



## The Toolbox

Microsoft Visual Studio  
Tableau  
R

## A Look at Individual ZIP Codes



## Predicting Discrete Amount (AMNTLVL) Using Entropy

**Predicting Discrete Amount (AMNTLVL) Using Entropy**

```

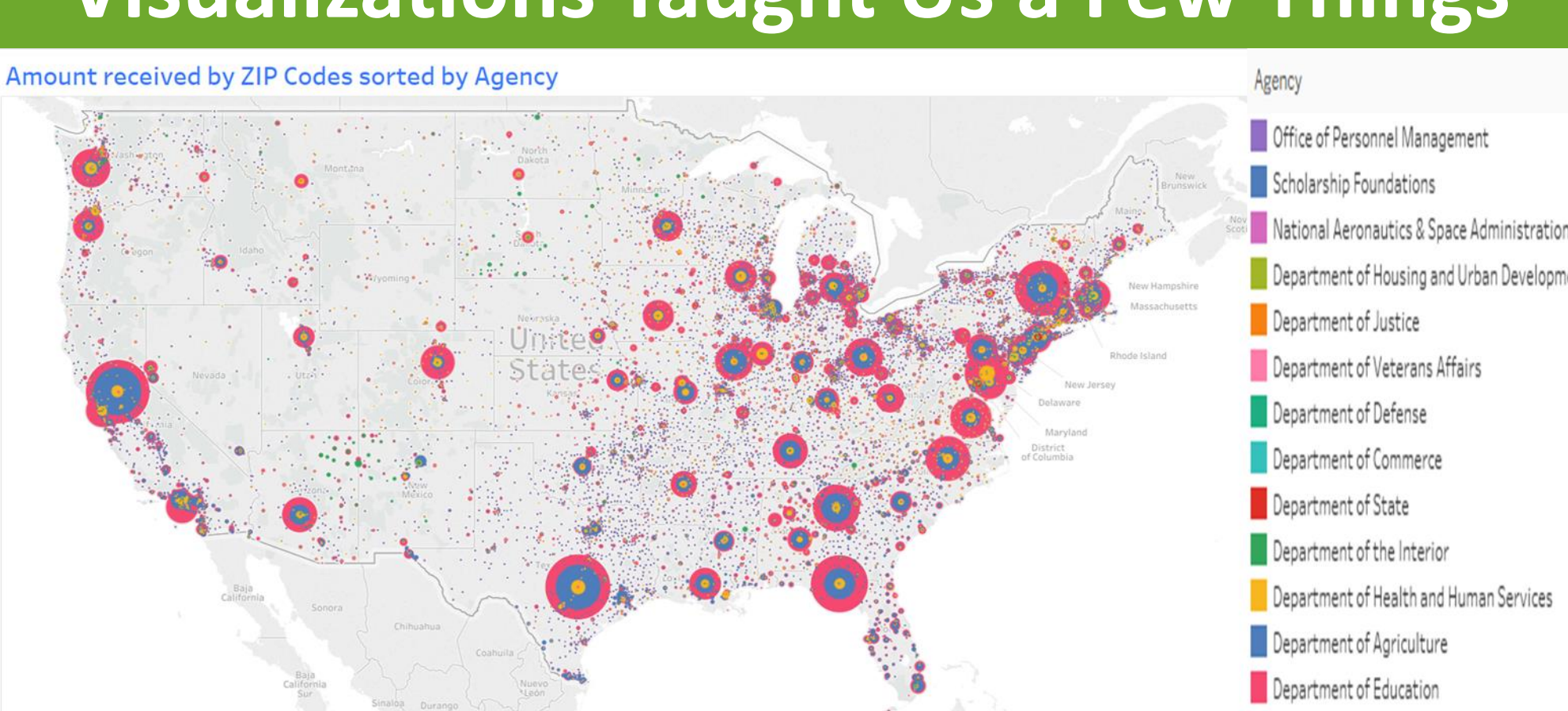
Mining model structure:
AGENCY
AMNTLVL
CATEGORY
CFDAMAJOR
OVERSIGHTAGENCY
PER_0_10
PER_10_19
PER_HS_GRAD
PER_POVERTY
PER_WHITE
UID
    
```

## Predicting Department when

```

ffs <- subset(ff, ff$PER_0_10 <= 1.643)
& (ff$PER_10_19 >= 24.28)
& (ff$PER_POVERTY <= 35.823)
ff$AGENCY <- ctree(data=ffs, as.factor(AGENCY_SHORT)~
PER_WHITE+PER_POVERTY+
PER_HS_GRAD+PER_BACH_DEGREE+
PER_0_10+PER_10_19)
    
```

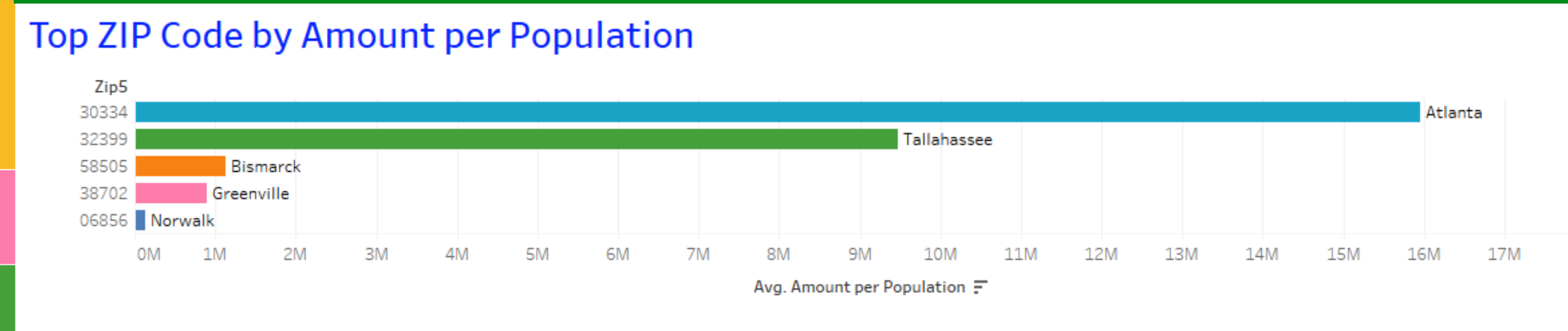
## Visualizations Taught Us a Few Things



## Baltimore 21218

City	Zip5	Amount
Tallahassee	32399	2,489,730,295.00
Sacramento	95814	3,647,477,841.00
Raleigh	27609	1,491,055,076.00
Baltimore	21218	6,551,005,703.00
Austin	78774	3,441,915,957.00
Atlanta	30334	1,698,657,059.00
Albany	12224	2,276,244,000.00

## Atlanta 30334 & Tallahassee 32399



## Decision Tree Using Bayesian Dirichlet Equivalent with Uniform Prior

**Decision Tree Using Bayesian Dirichlet Equivalent with Uniform Prior**

```

Mining Legend
Value Cases Probability Histogram
Agriculture/Food 9634 1.23%
Education 30240 38.45%
Environment 37345 4.73%
Healthcare 28146 35.39%
Housing 90403 11.49%
Missing 0 0.00%
Small Business 3923 0.50%
Transportation 43720 5.56%
    
```

## Predicting Department when

```

ffs <- subset(ff, ff$PER_0_10 <= 1.643)
& (ff$PER_10_19 >= 24.28)
& (ff$PER_POVERTY <= 35.823)
ff$AGENCY <- ctree(data=ffs,
as.factor(AGENCY_SHORT)~
PER_WHITE+PER_POVERTY+
PER_HS_GRAD+PER_BACH_DEGREE+
PER_0_10+PER_10_19)
    
```

## Education Funds Granted by Dept. of Education (in dollars)

FEDERAL PELL GRANT PROGRAM 36,983,218,092	SPECIAL EDUCATION - GRANTS TO STATES 24,677,069,616	IMPROVING TEACHER QUALITY STATE GRANTS 16,296,010,396
FEDERAL DIRECT STUDENT LOANS 25,255,230,199	SPECIAL EDUCATION GRANTS TO STATES WILLIAM D. FORD FEDERAL TWENTY-FIRST CENTURY IMPACT AID	

## Los Angeles 90071

**OUR STUDENTS**

- 98% of Students are Eligible for Free or Reduced Price Meals
- 90% of Students are English Learners
- 24% of Students are English Learners

**EIN 261607268**

Federalprogramname

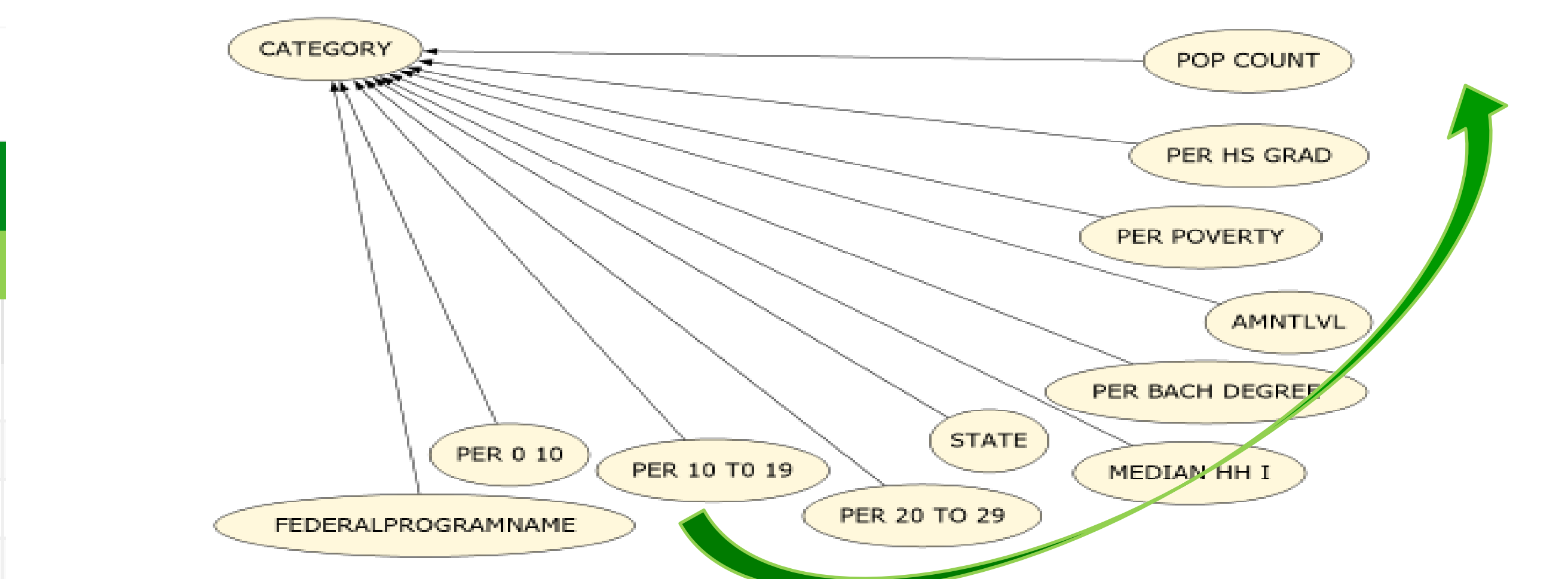
TITLE II PART A IMPROVING TEACHER QUALITY

TITLE III PART A LIMITED ENGLISH PROFICIENCY

TITLE V PART B PUBLIC CHARTER SCHOOLS GRANTS

CHILD NUTRITION PROGRAMS

## Cluster Analysis for Predicting Category Top Features



## Prediction Models & Lift Charts

**Prediction Models & Lift Charts**

Microsoft Decision Trees

Structure

- AGENCY
- AMNTLVL
- CATEGORY
- MEDIAN HH I
- PER\_0\_10
- PER\_10\_19
- PER\_HS\_GRAD
- PER\_POVERTY
- PER\_WHITE
- UID