Information Graphics for Actuaries

Garrett Bradford / Cody Webb

March 15, 2016

Outline

- What is data visualization?

- Picking your platform

- Introduction to GIS

- Principles of data visualization

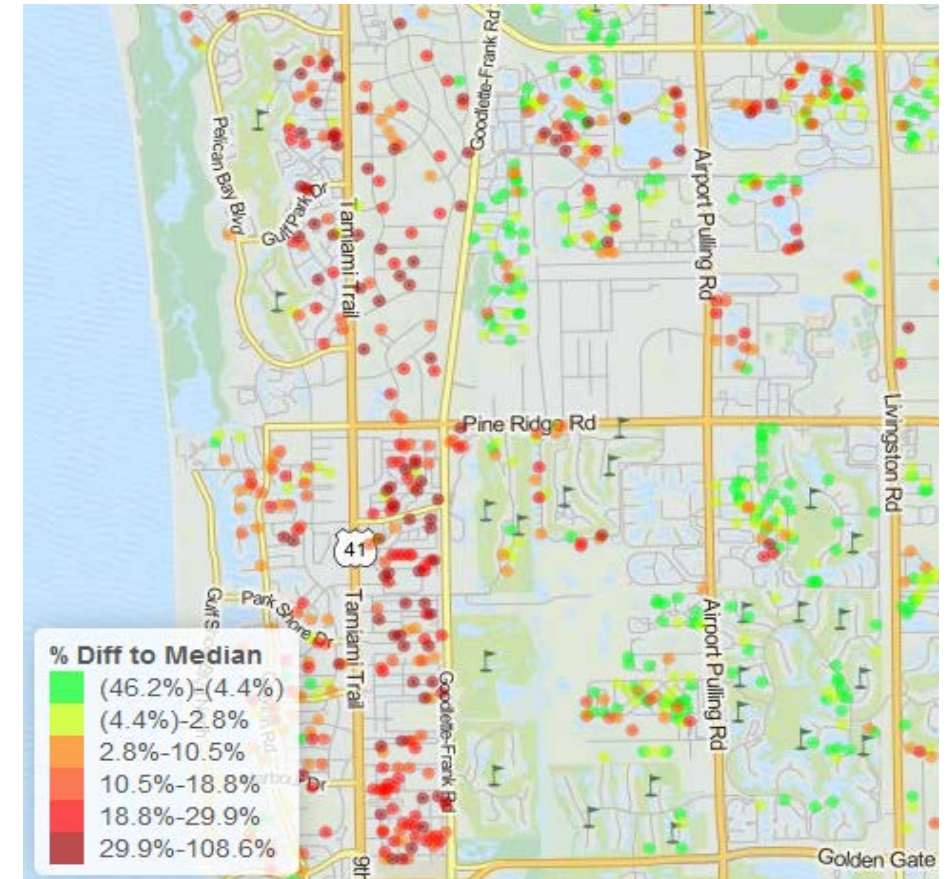- Guide to creating graphics

- Case Study: Milliman Pixel

Milliman

Data Visualization

Graphical representation of data

- Tables
- Bar charts
- Scatter plots

Data visualization is both an art and a science

Lon	Lat	Diff
30.40686	-84.2982	1.052
30.40666	-84.2971	0.688
30.40686	-84.2975	0.331
30.40766	-84.2975	0.204
30.40648	-84.2966	0.256
30.40667	-84.2966	0.568
30.40806	-84.2979	-0.418
30.40634	-84.2963	0.091
30.40634	-84.2956	0.812
…	…	….

% Diff to Median
- (46.2%)-(4.4%)
- (4.4%)-2.8%
- 2.8%-10.5%
- 10.5%-18.8%
- 18.8%-29.9%
- 29.9%-108.6%

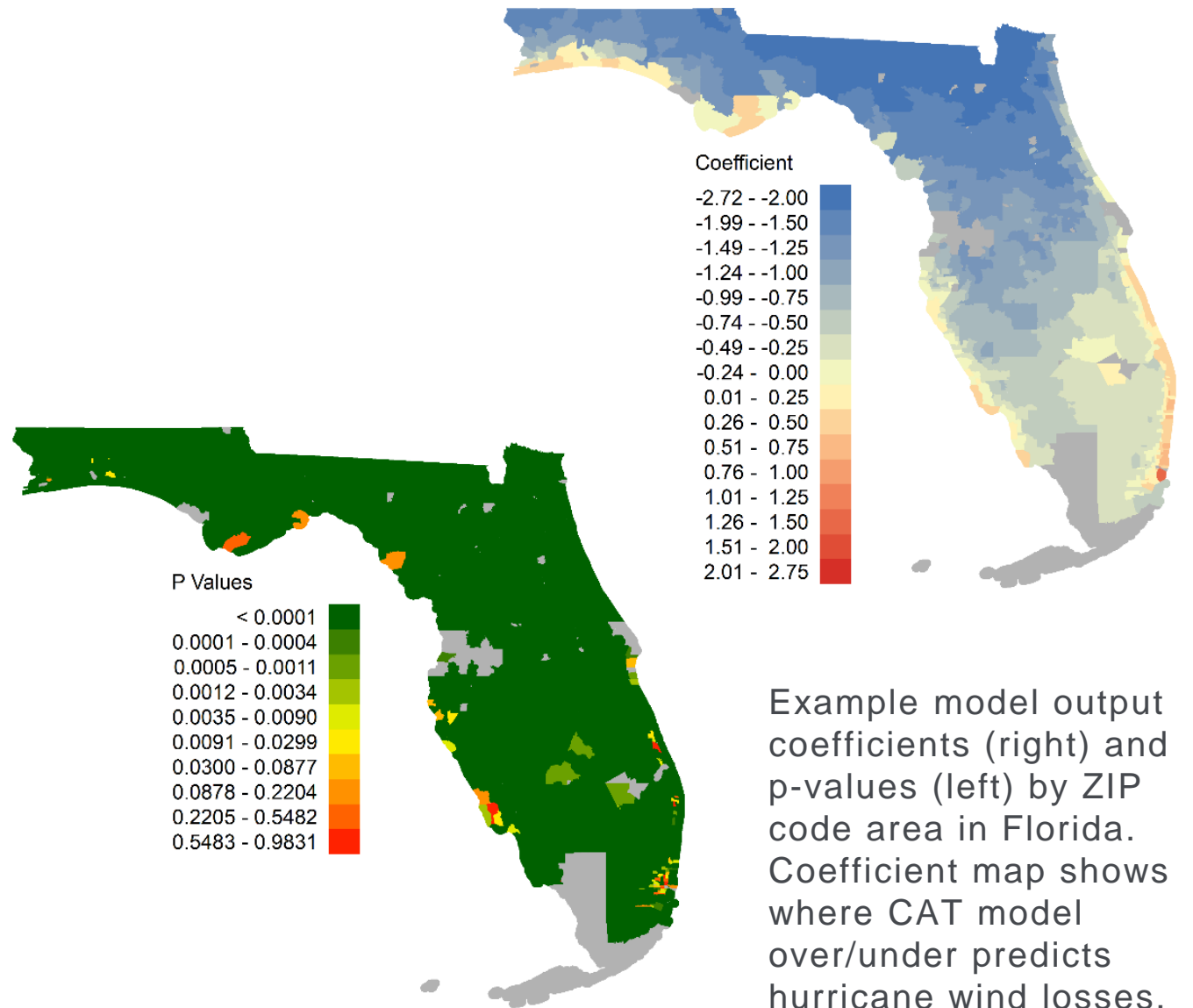SOURCE: Pixel
http://milliman.com/pixel

Milliman

Data Visualization

Just a substitute for tables?

- Efficiency

- New perspectives

Used in every step of the analysis

- Data exploration

- Data modeling/review

- Validation of results

- Presentation of findings

Coefficient

- -2.72 - -2.00
- -1.99 - -1.50
- -1.49 - -1.25
- -1.24 - -1.00
- -0.99 - -0.75
- -0.74 - -0.50
- -0.49 - -0.25
- -0.24 - 0.00
- 0.01 - 0.25
- 0.26 - 0.50
- 0.51 - 0.75
- 0.76 - 1.00
- 1.01 - 1.25
- 1.26 - 1.50
- 1.51 - 2.00
- 2.01 - 2.75

P Values

- < 0.0001
- 0.0001 - 0.0004
- 0.0005 - 0.0011
- 0.0012 - 0.0034
- 0.0035 - 0.0090
- 0.0091 - 0.0299
- 0.0300 - 0.0877
- 0.0878 - 0.2204
- 0.2205 - 0.5482
- 0.5483 - 0.9831

Example model output coefficients (right) and p-values (left) by ZIP code area in Florida. Coefficient map shows where CAT model over/under predicts hurricane wind losses.

Milliman

Finding the Right Tool

- Desktop Applications

- Excel, ArcMap, Illustrator

- Coded solutions

- Matlab, R, SAS, Python

- Web Applications

- Tableau, Google Chart Tools, CartoDB
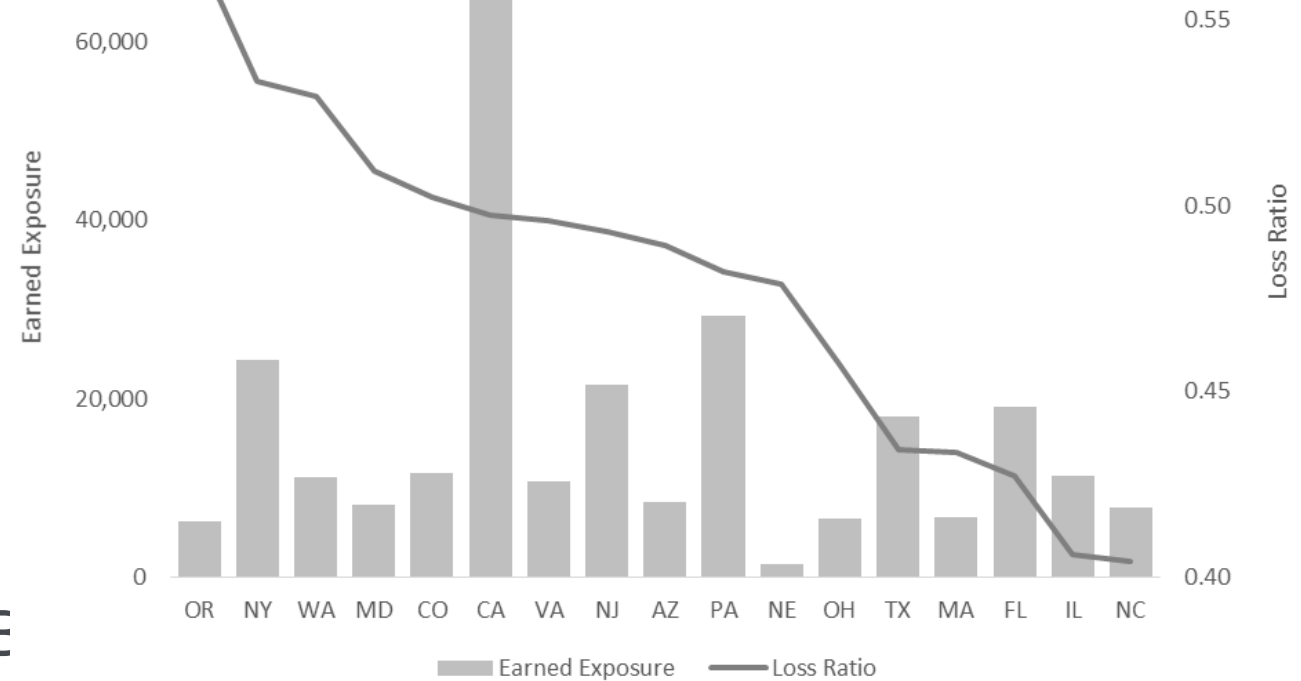
- Milliman Pixel


```
proc sgplot data=input;
where pettype='Dog';
vbar state/response=EarnedExposure;
vline  state/response=lr y2axis;
run;
```

Introduction to GIS

Geographic information systems (GIS)

GIS is used to capture, store, analyze, and present geographic data

| OBJECTID *	county *	Owner_SF_detached	policies	quotes
1	ABBEVILLE	5859	1	19
2	AIKEN	35926	37	174
3	ALLENDALE	1610	1	14
4	ANDERSON	43918	30	141
5	BAMBERG	3125	7	56
6	BARNWELL	3911	8	29
7	BEAUFORT	36393	629	11633
8	BERKELEY	35246	198	3427
9	CALHOUN	3317	29	170
10	CHARLESTO	71596	743	11770
11	CHEROKEE	9926	146	571
12	CHESTER	7047	45	142
13	CHESTERFIE	8191	6	37
14	CLARENDON	5575	54	350
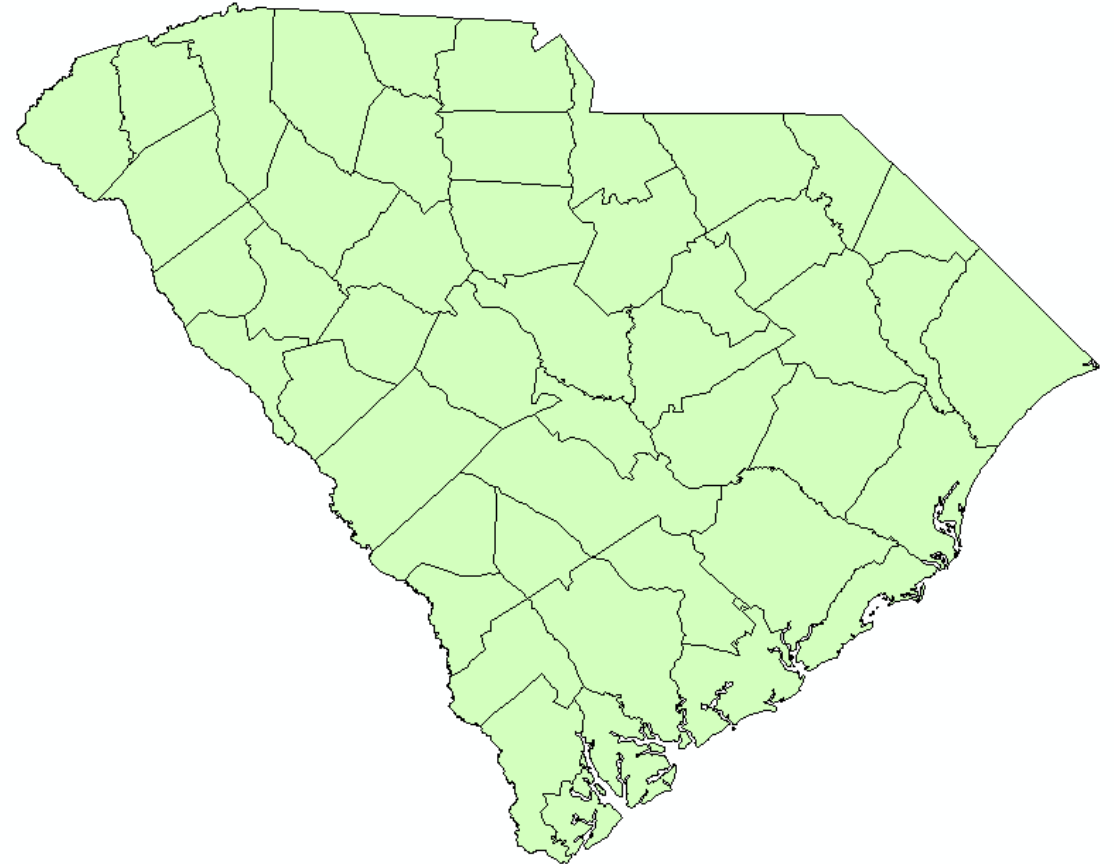
Combining actuarial data with GIS

Existing geographies:

- Territory
- Zip code
- County

Identifying new geographic characteristics:

- Geocoding
- Distance calculation
- Clustering
- Third-Party data
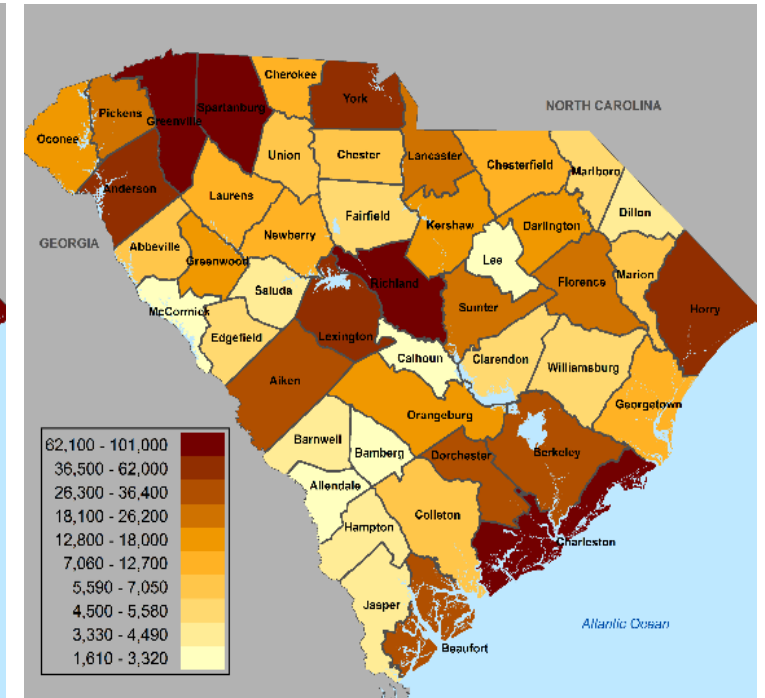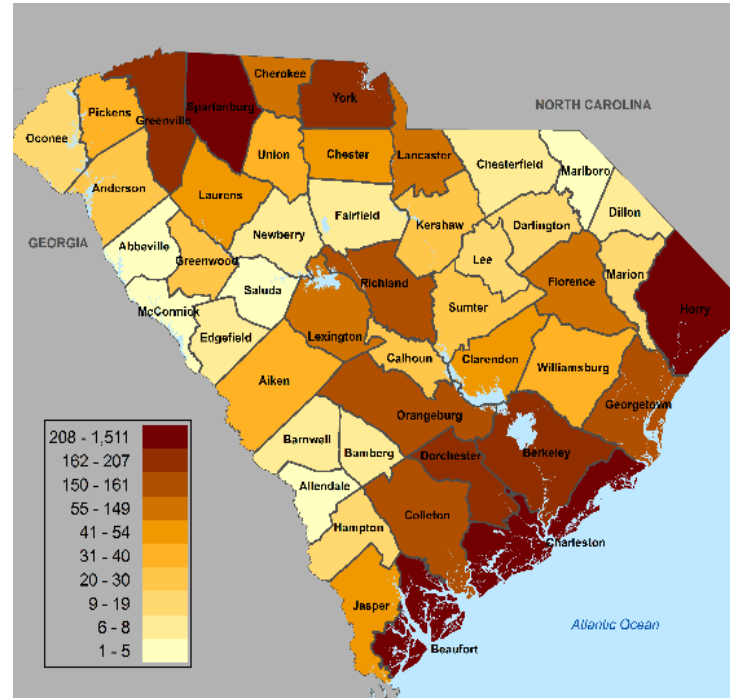
Comparison of example company's HO-3 in-force policy distribution (left) with estimate of total single family homes by county (right).

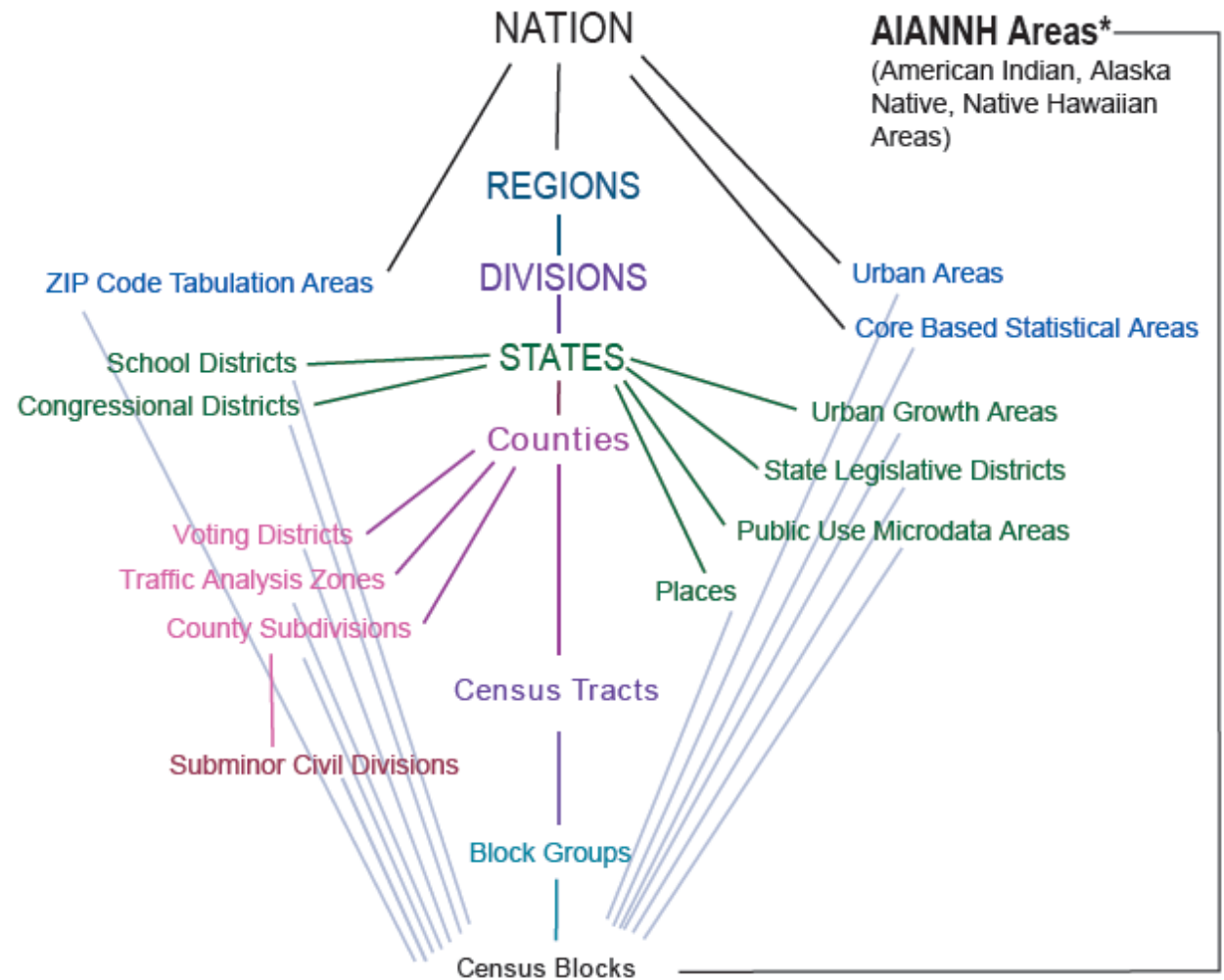# Mapping Solutions

Desktop GIS applications:

- ESRI's ArcMap, GRASS, QGIS
- R, SAS, SQL

Web applications:

- ArcGIS Online, CartoDB

Data sources:

- TIGER files from US Census
- National Atlas, USGS, NOAA, local government
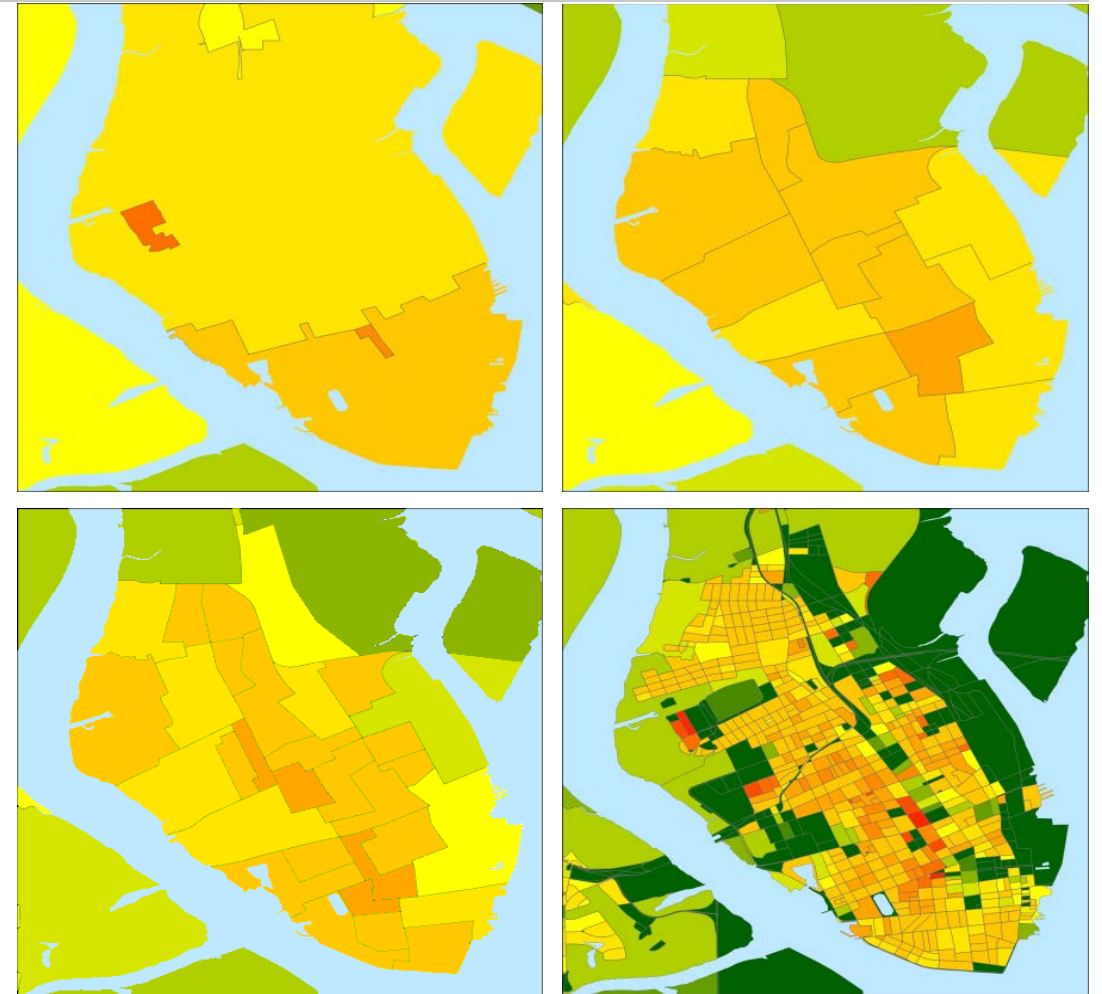

Source: U.S. Census Bureau
https://www.census.gov/geo/reference/

Mapping Considerations

What geography is right for your analysis?

Modifiable areal unit problem (MAUP):

"The areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating"
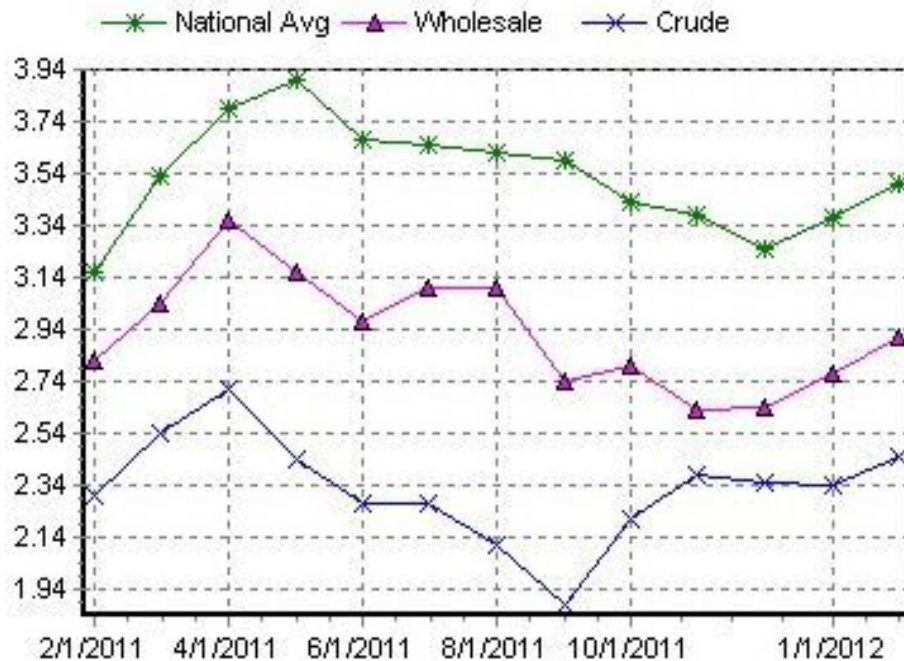
(Openshaw, 1983)

Population density in Charleston, SC plotted using ZIP Code Tabulation Areas (ZCTAs), Census Tracts, Block Groups, and Blocks.

Milliman

Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

SOURCE:
Fox News, America's Newsroom, 2/20/12
AAA Daily Fuel Gauge Report, 2/21/12
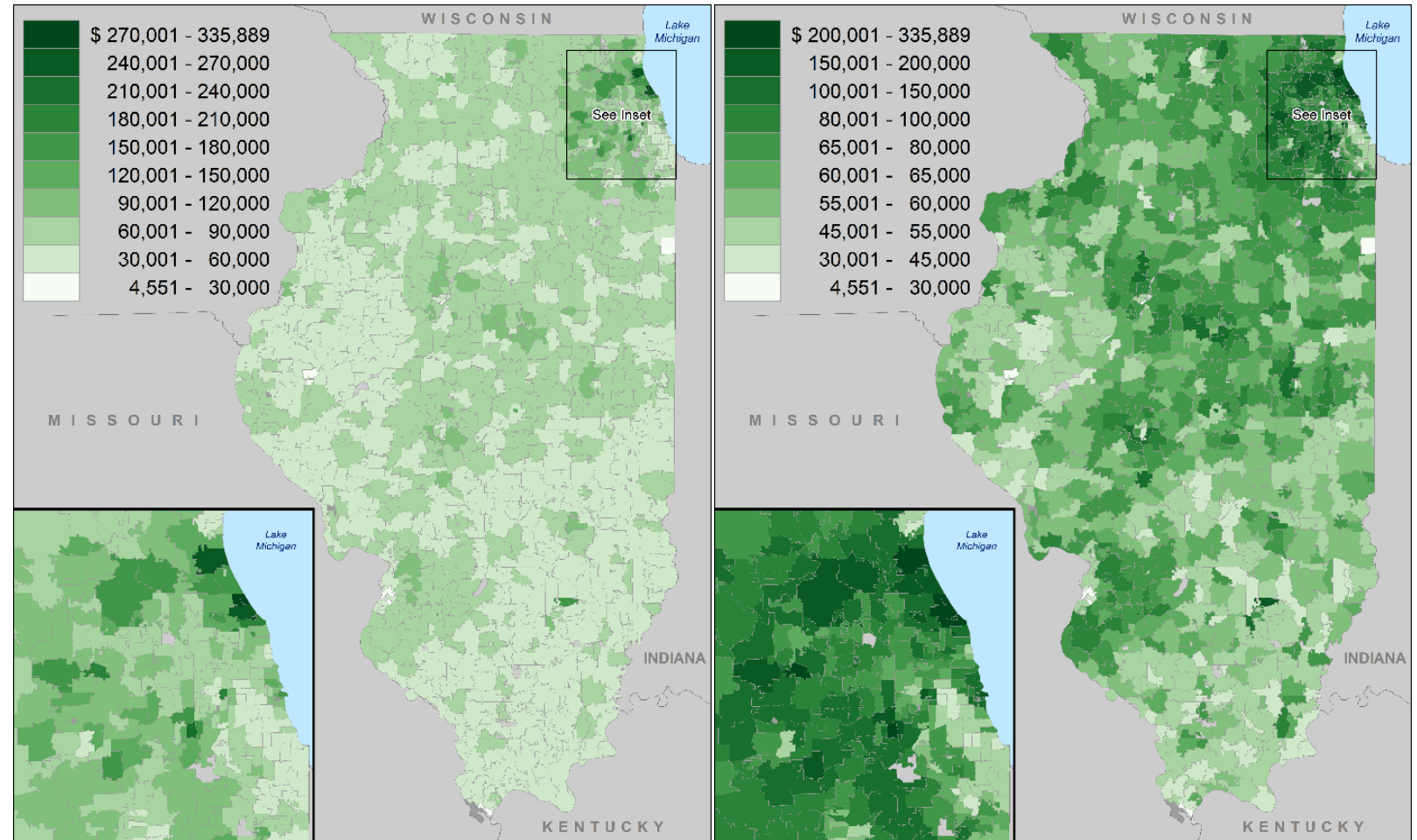via Media Matters
(www.mediamatters.org)

Milliman

Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

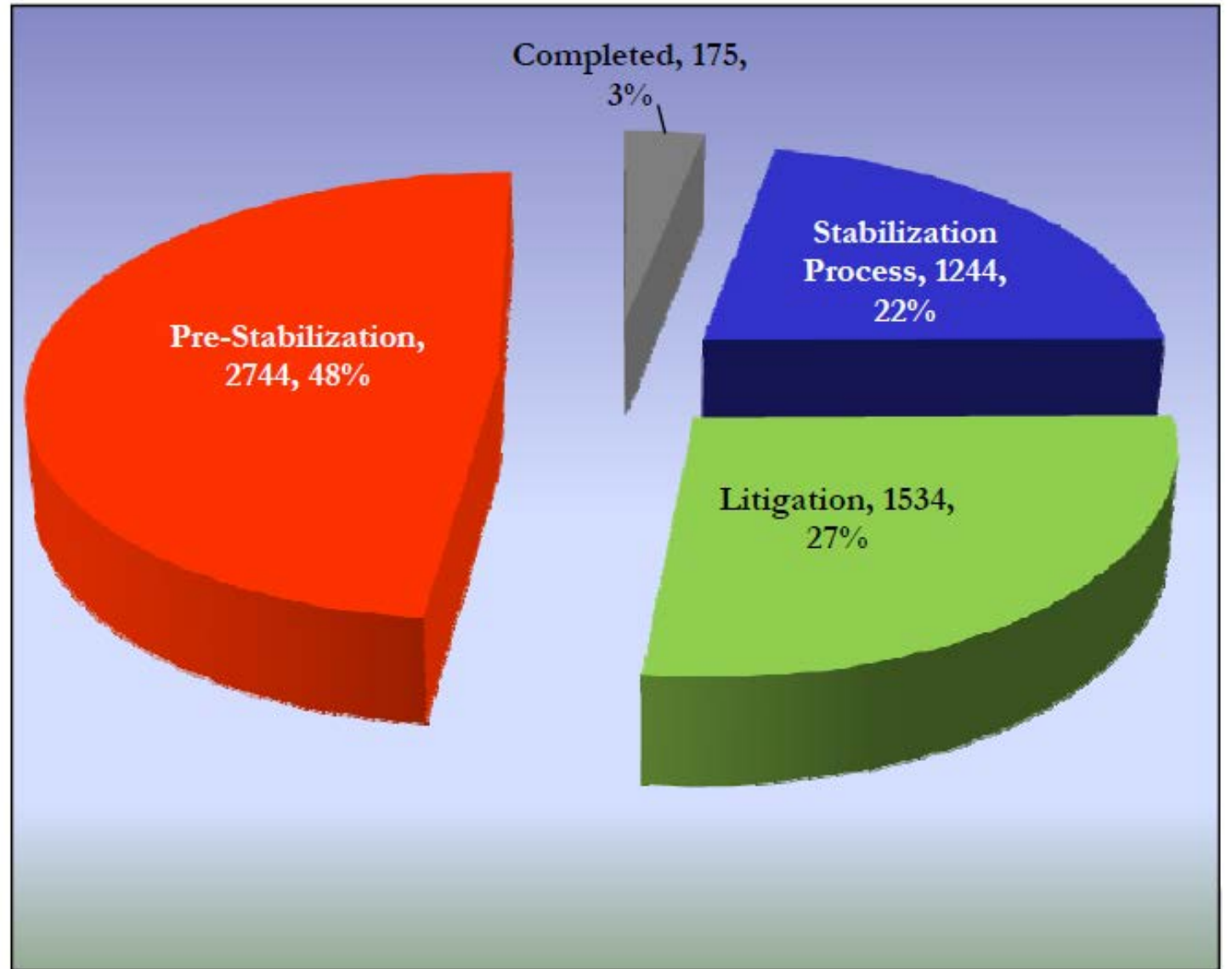## Median Household Income by ZCTA

Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

SOURCE: "Distribution of Sinkhole Pending Inventory" Florida Rate Filing #14-17638
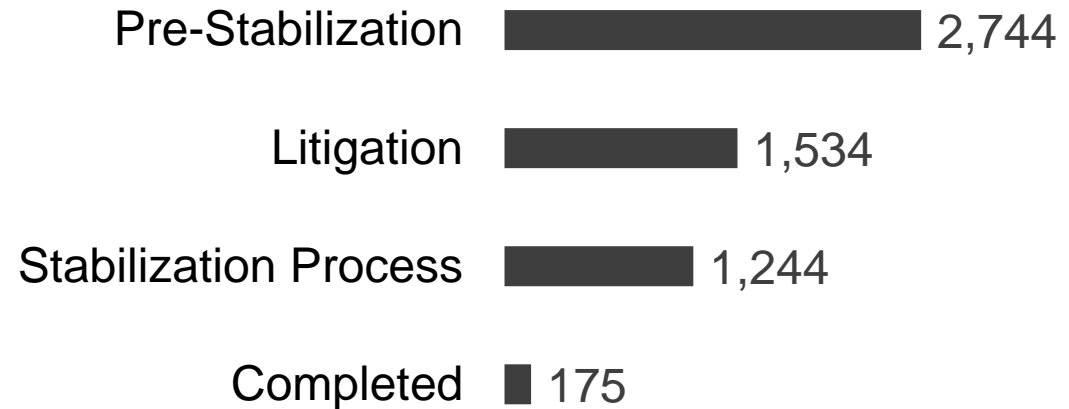
Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

Type	Claims	%
Pre-Stabilization	2,744	48%
Litigation	1,534	27%
Stabilization Process	1,244	22%
Completed	175	3%

Sinkhole Claims

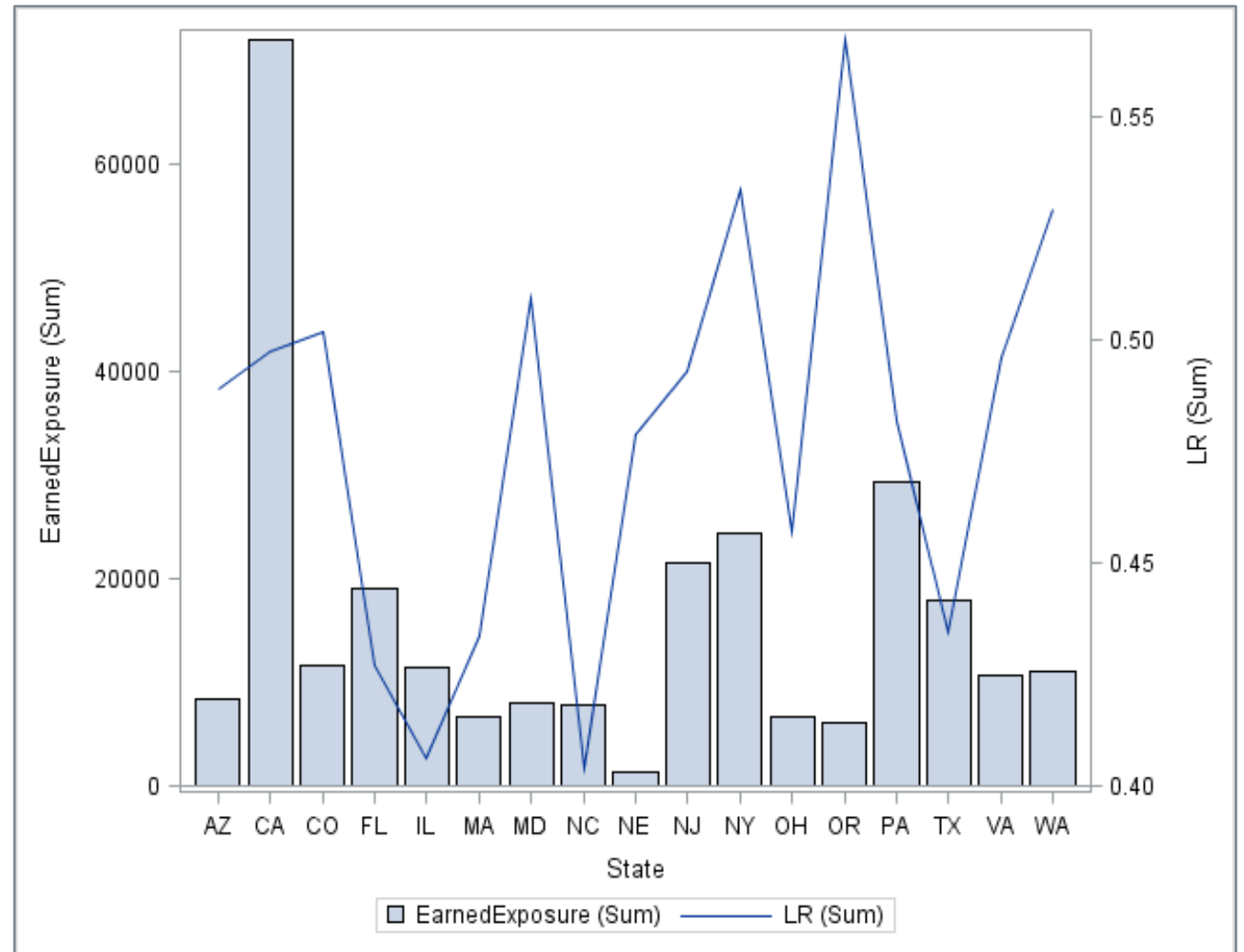|	
Pre-Stabilization	2,744
Litigation	1,534
Stabilization Process	1,244
Completed	175

Milliman

Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

```
proc sgplot data=input;
where pettype='Dog';
vbar state/response=EarnedExposure;
vline  state/response=lr y2axis;
run;
```
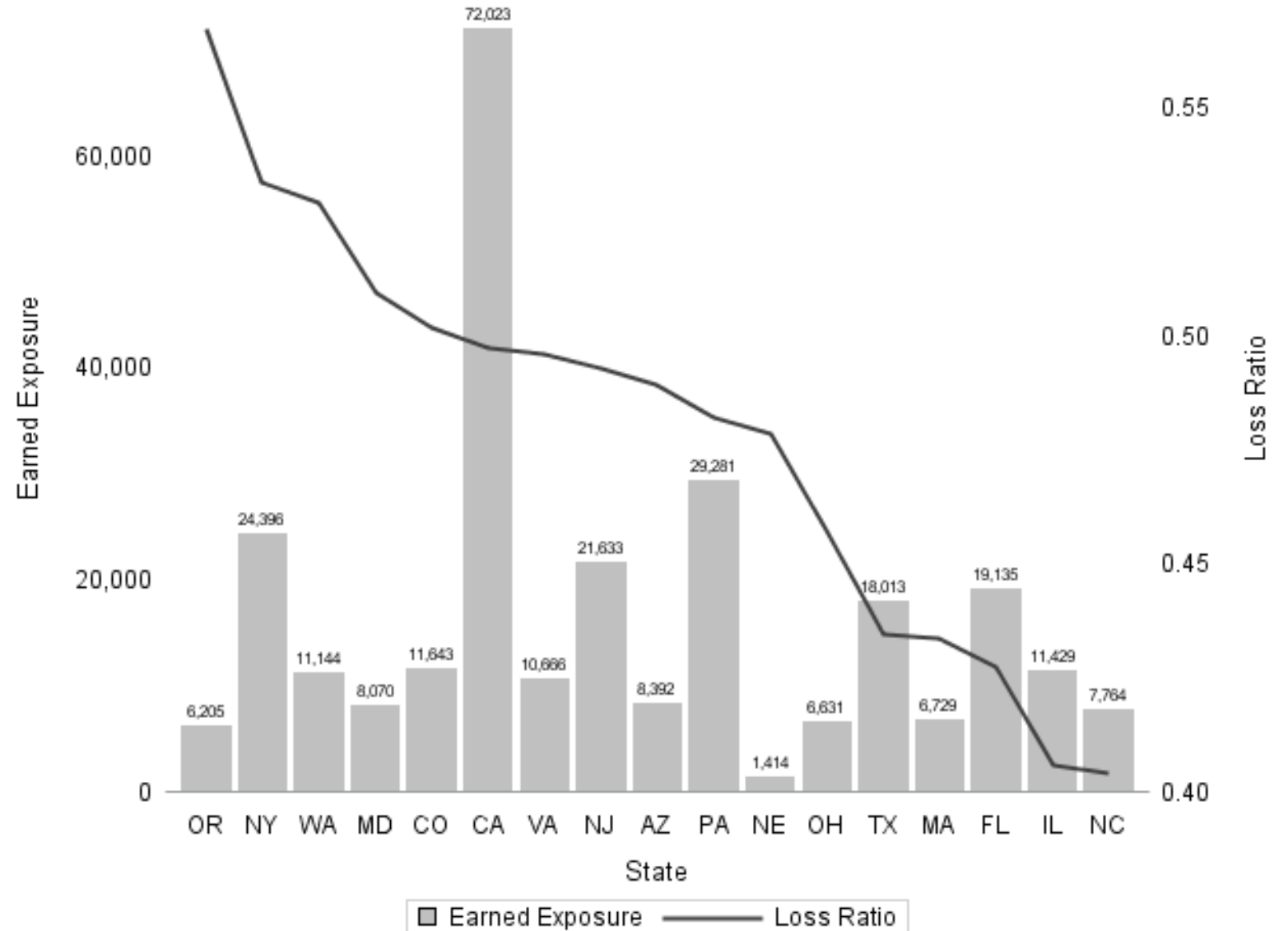
Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

Basic Principles

Honesty

Remove junk / ink

Revise, revise, revise

```
proc sgplot data=input NOBORDER NOWALL;*No need for outlines;
where pettype='Dog';
vbar state / response=EarnedExposure
  legendlabel='Earned Exposure'      *Clear labels;
  datalabel=EarnedExposure           *Directly label data;
  DATALABELATTRS=(Size=6)            *> detail, > clarity;
  NOOUTLINE
  FILLATTRS=(COLOR=GRAYC0);          *No need for color;
vline  state / response=lr
  y2axis
  legendlabel='Loss Ratio'
  categoryorder=respdesc
  lineattrs=(THICKNESS= 2 PT COLOR=GRAY40);
yaxis
  label='Earned Exposure'
  DISPLAY=(NOTICKS NOLINE);
y2axis
  label='Loss Ratio'
  DISPLAY=(NOTICKS NOLINE);
xaxis
  label='State'
  DISPLAY=(NOTICKS NOLINE);
format EarnedExposure comma8.0 lr 6.2; *Format numbers;
run;
```
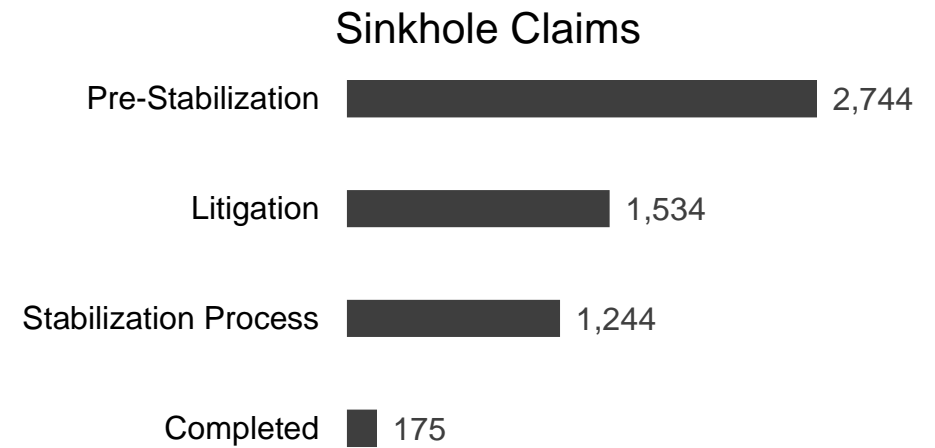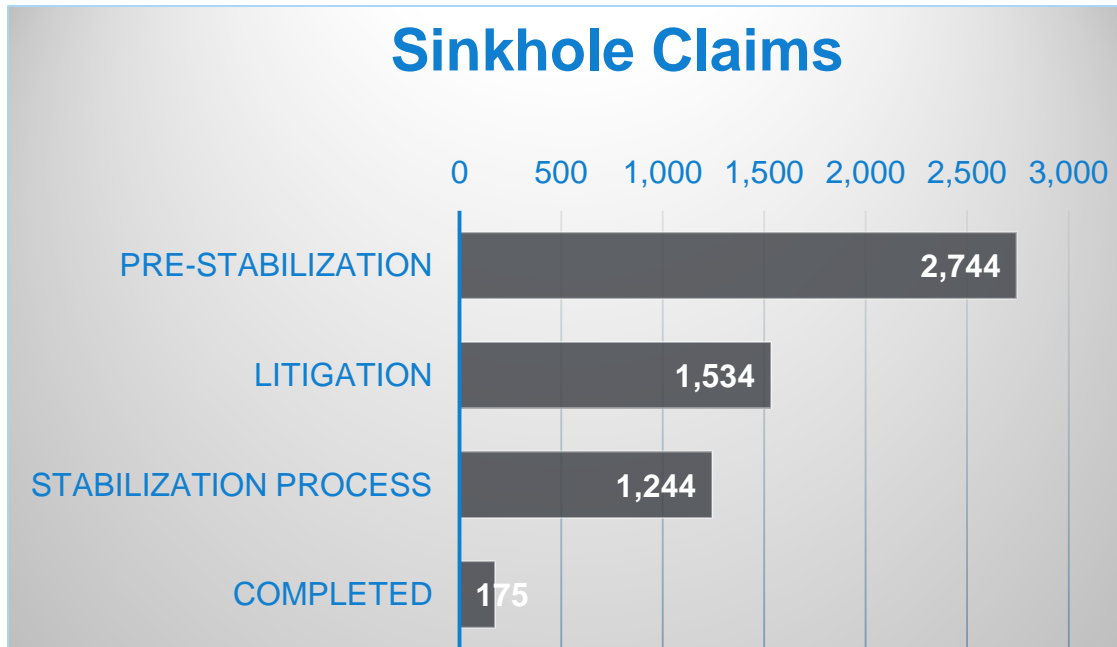
Guide to Creating Graphics

Do's

- Heading 2pt > body
- Not too small
- Simple font
- Data range = 2/3 y-axis
- Direct labeling
- Sort data in meaningful way
- Max 3-4 lines per chart

Don'ts

- Heavy gridlines
- 3D bars
- Pie charts
- Let type oppress graphics
- No ALL CAPS or bold / italic
- Multi-color for no reason
- Awkward increments (3,6,9…)

Adapted from Wong, Dona. *The Wall Street Journal Guide to Information Graphics: The Do's and Don't of Presenting Data, Facts, and Figures*.

Milliman

Avoid defaults

Guide to Creating Graphics

- Show your data
- Induce thinking about substance
- Avoid distortion
- Many numbers/small space
- Make large datasets coherent
- Encourage eye-to-eye comparison
- Show multiple levels of detail
- Have a clear purpose
- Integrate statistical and verbal descriptions of data

Adopted from Tufte, Edward R.
*The Visual Display of Quantitative Information.* 2nd Edition.

Milliman
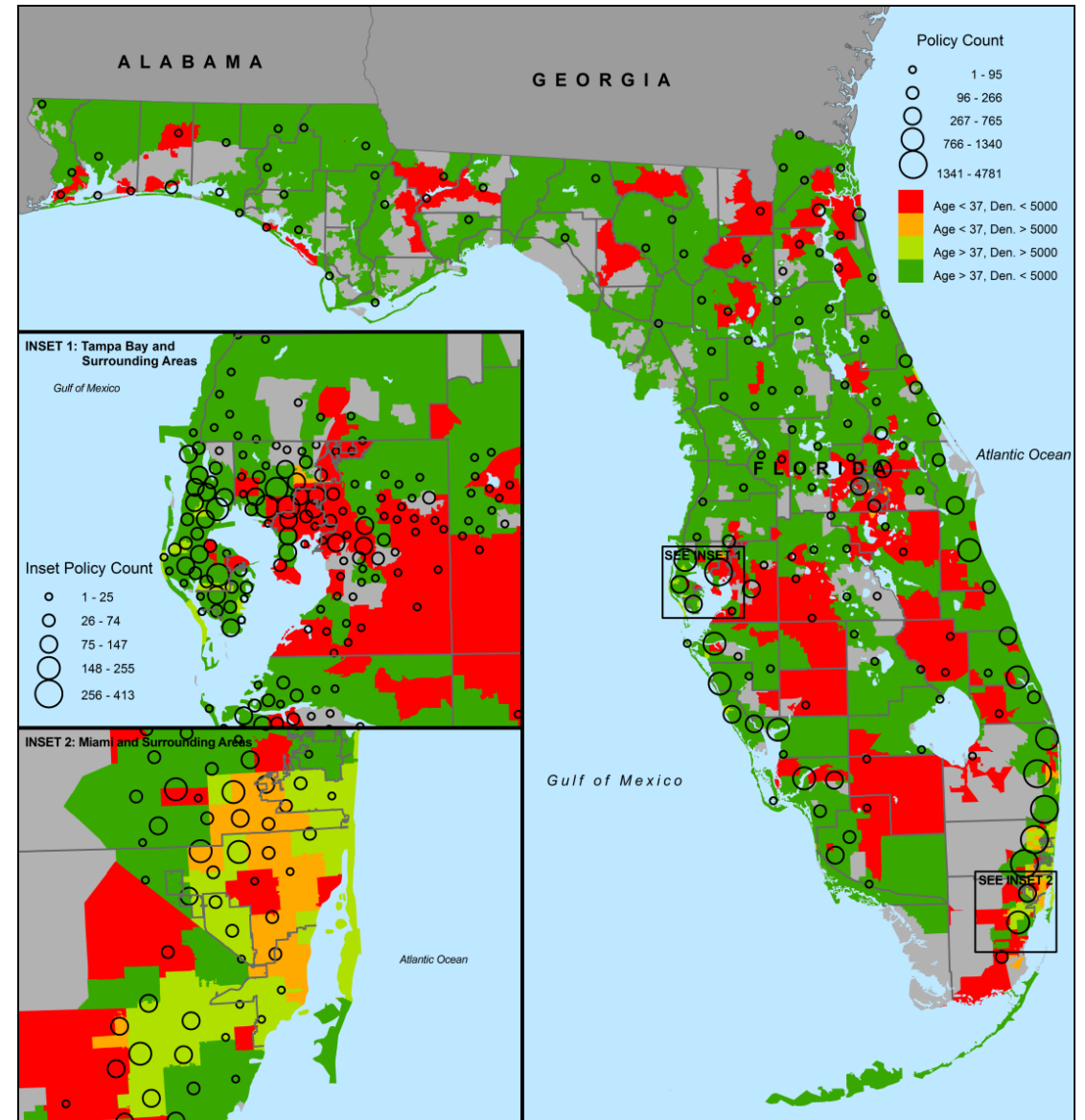
Case Study: Pixel

Pixel

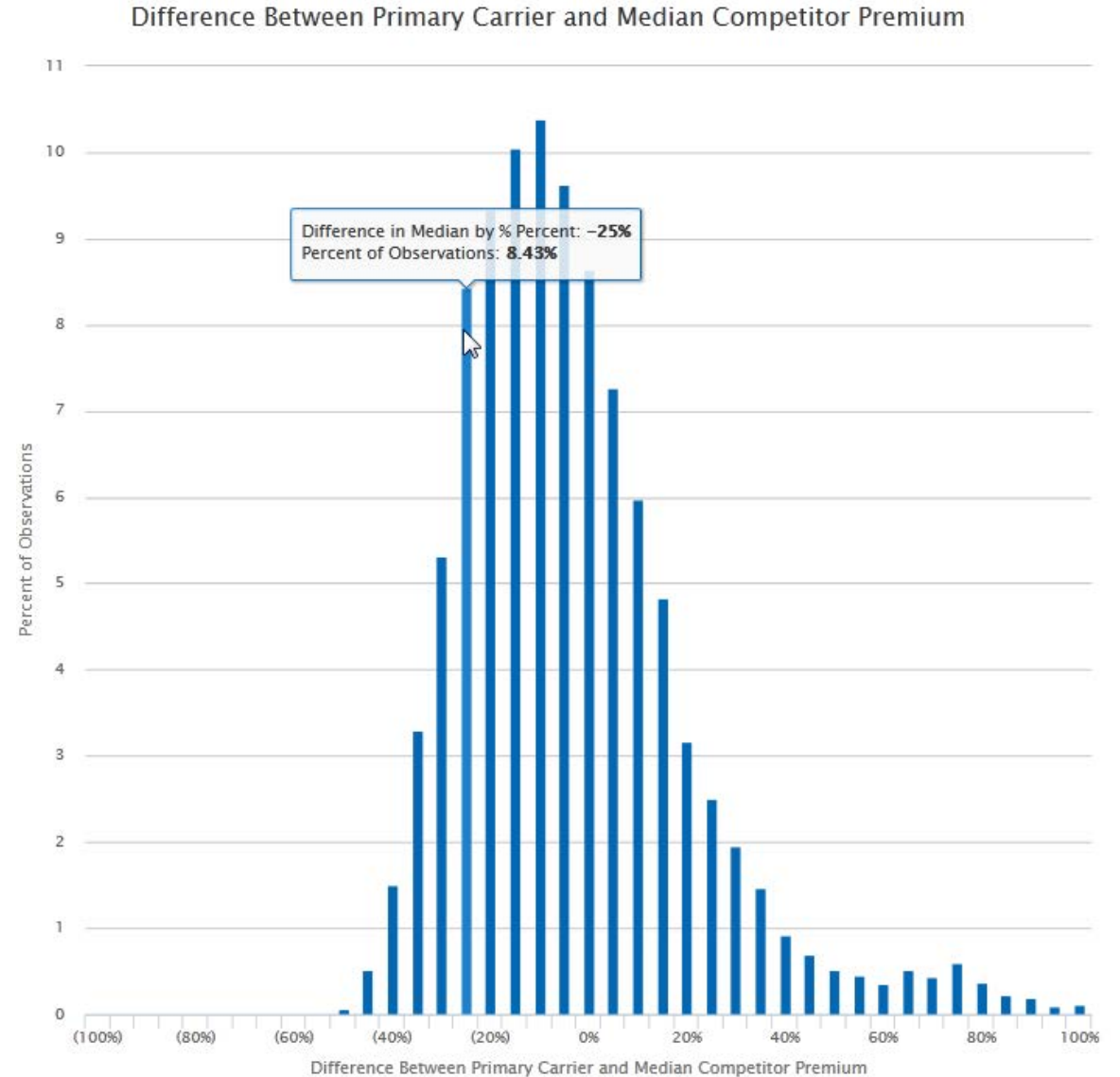Online tool designed to explore your company's competitive position

- Who are your most significant competitors?

- What variables/segments might be mispriced versus the market?

- Where do you want to make competitive price adjustments, and how much should they be?

- What risk characteristics and/or segments should you target?
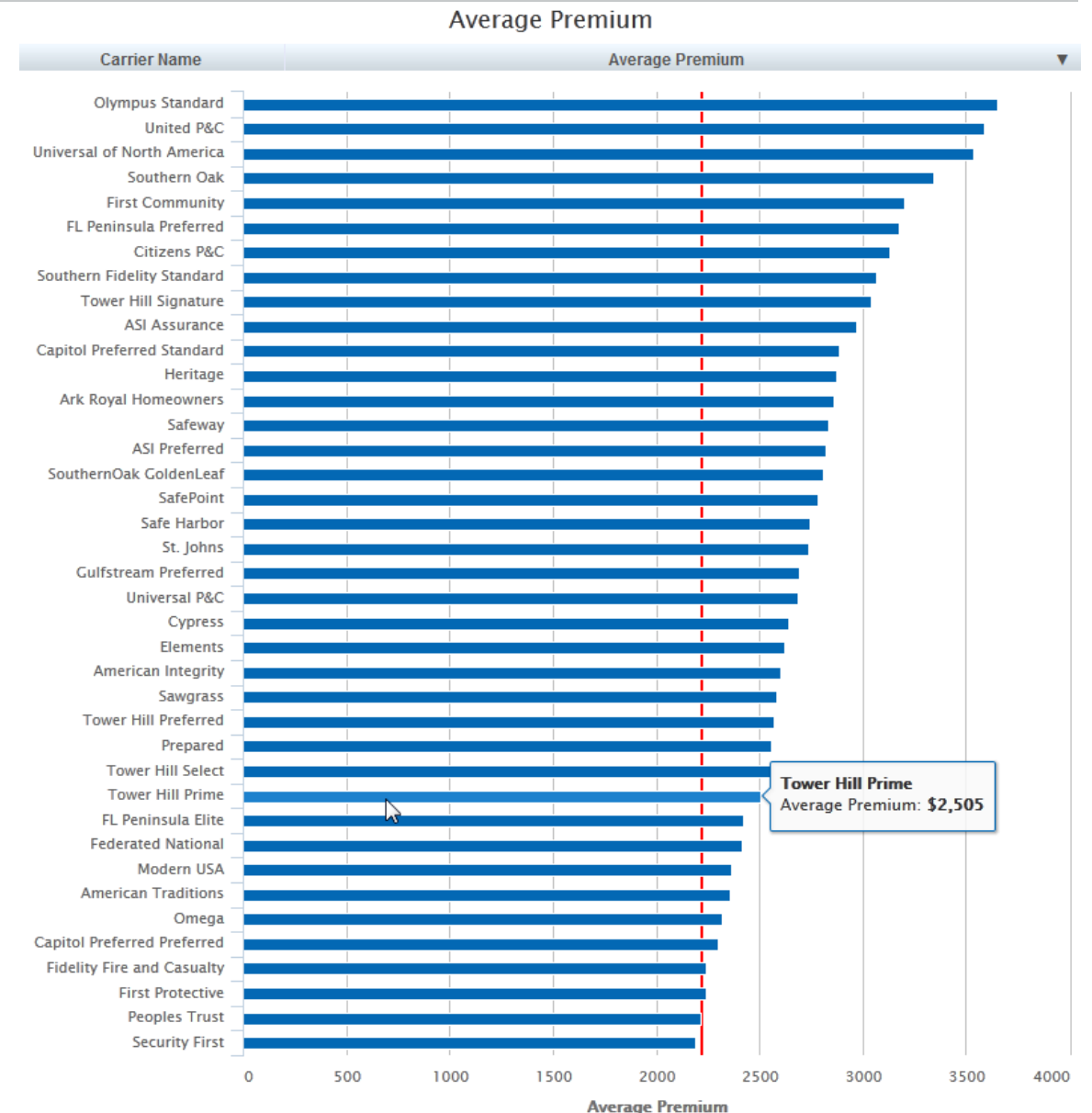
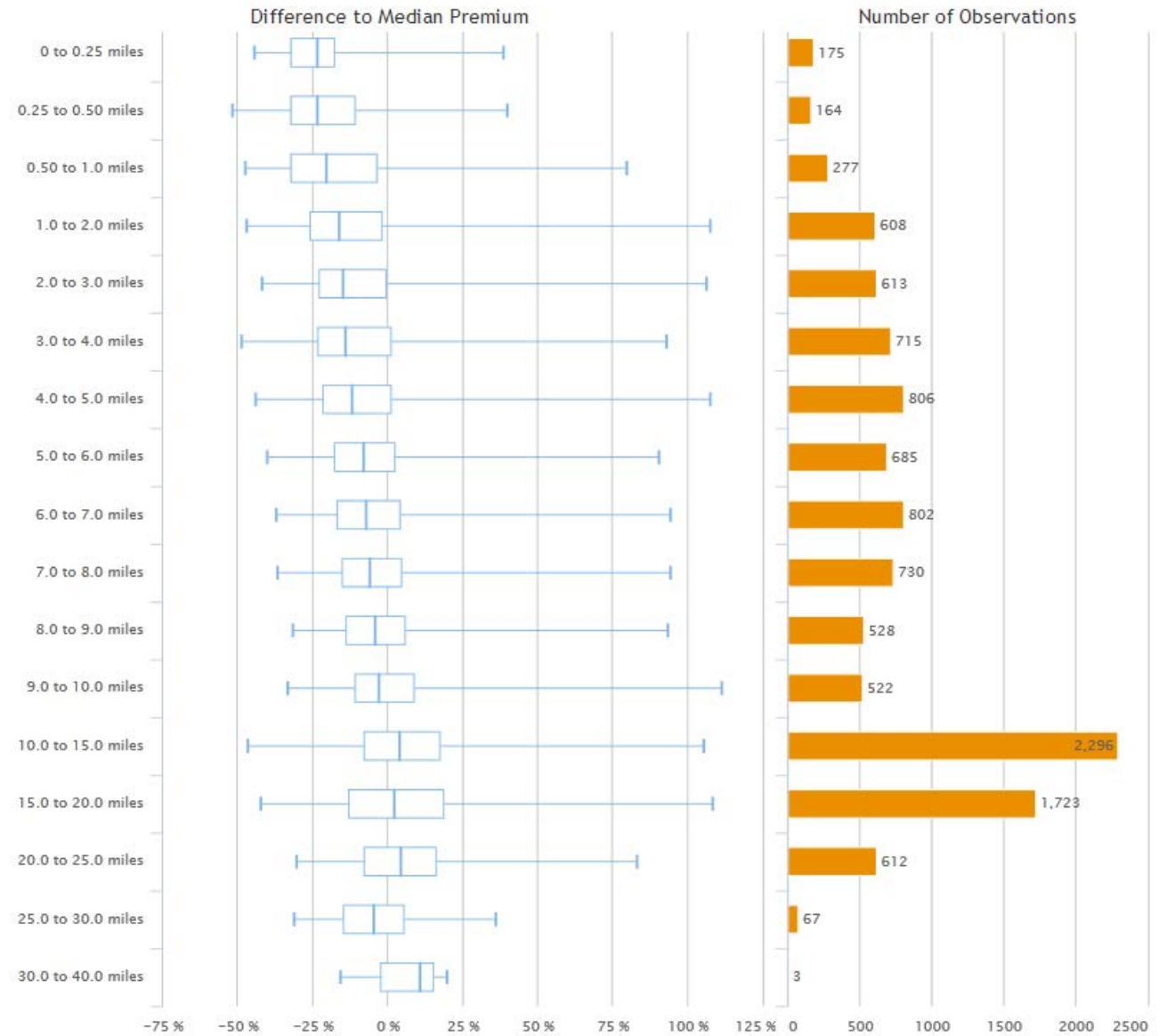# Pixel

How do I compare to the market?

▪ View percent or dollar difference to competitor premiums

Difference Between Primary Carrier and Median Competitor Premium

Difference in Median by % Percent: −25%
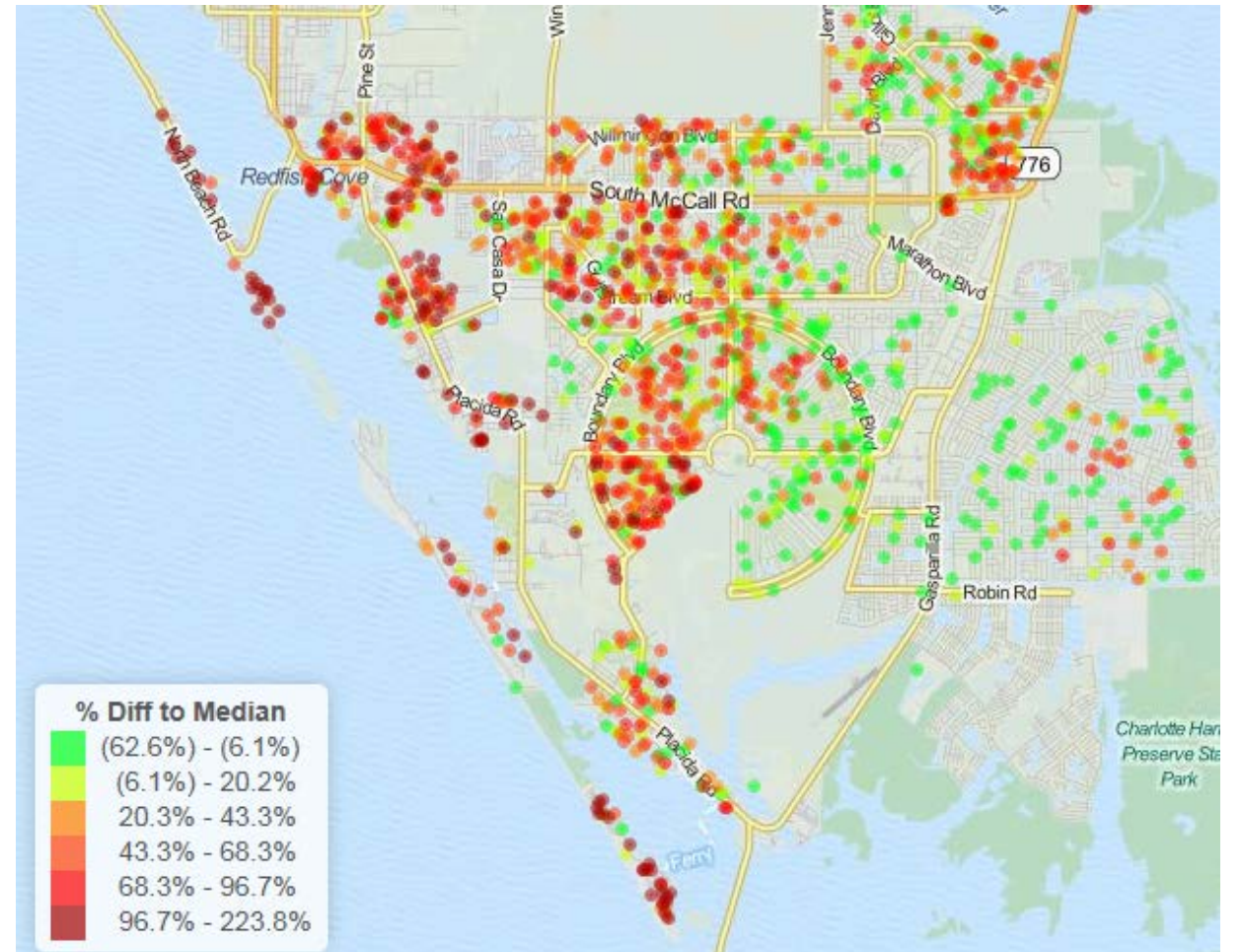Percent of Observations: 8.43%

Milliman

Pixel

Where am I most/least competitive?

▪ Summarize by rating variable

Pixel

Where am I most/least competitive?

- Map by county, territory, zip code, or policy

% Diff to Median

Color	Range
Green	(62.6%) - (6.1%)
Yellow-green	(6.1%) - 20.2%
Orange	20.3% - 43.3%
Orange-red	43.3% - 68.3%
Red	68.3% - 96.7%
Dark red	96.7% - 223.8%

Milliman

Questions?

Garrett Bradford
Cartographer / GIS Analyst
Milliman, Inc.
650 California Street, 17th Floor
San Francisco, CA 94108
garrett.bradford@milliman.com
(415) 394-3792

Cody Webb
Actuary
Milliman, Inc.
650 California Street, 17th Floor
San Francisco, CA 94108
cody.webb@milliman.com
(415) 394-3763

Thank you

March 15, 2016