Visualize it!
Data and Analytics
Visualization in Insurance

Wind Days Per Year (1995–1999)
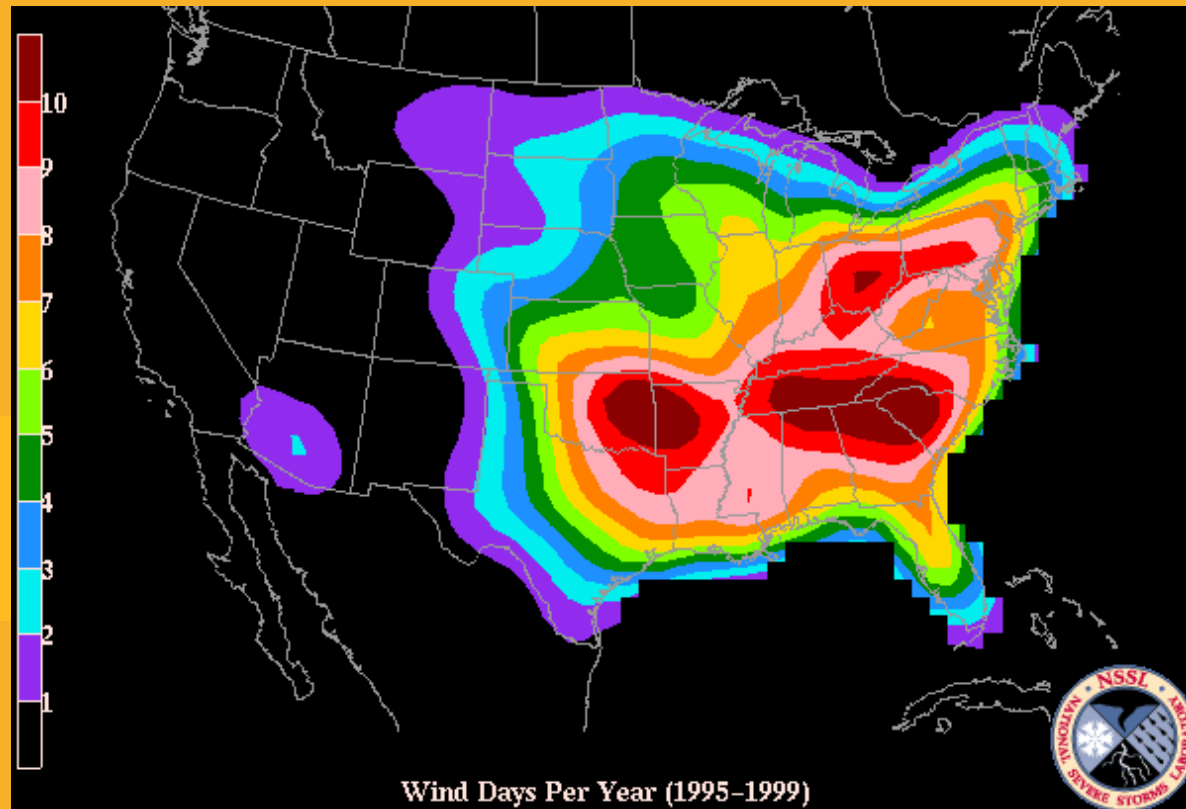
June 6, 2016
CARe, Boston, MA

Antitrust notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Effective communication?

Average HO-3 Rates by Territory within Region
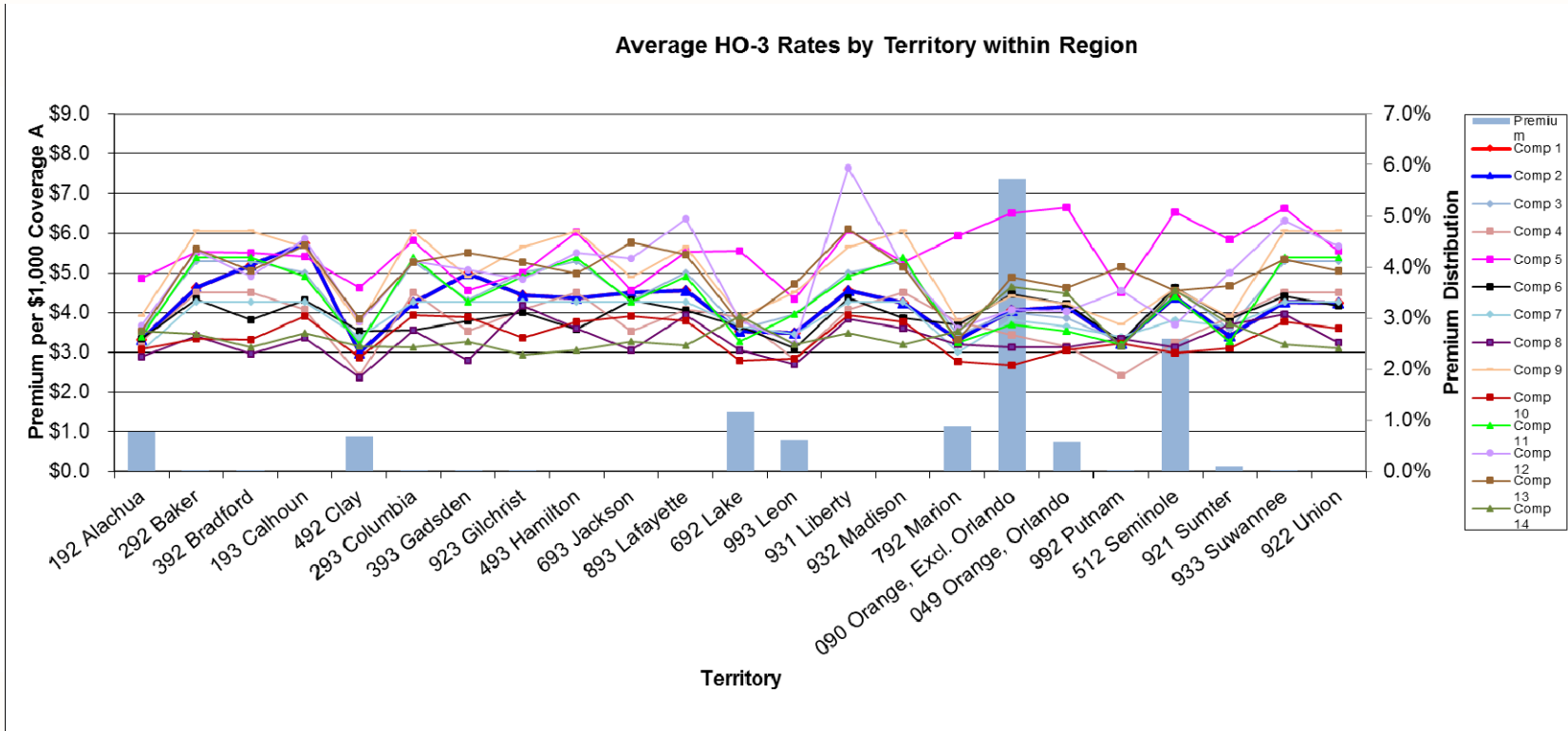
Speakers

Alice Underwood, Executive Vice President

WillisRe

Garrett Bradford, Cartographer/GIS Analyst

Milliman

Outline

- Data Visualization Basics

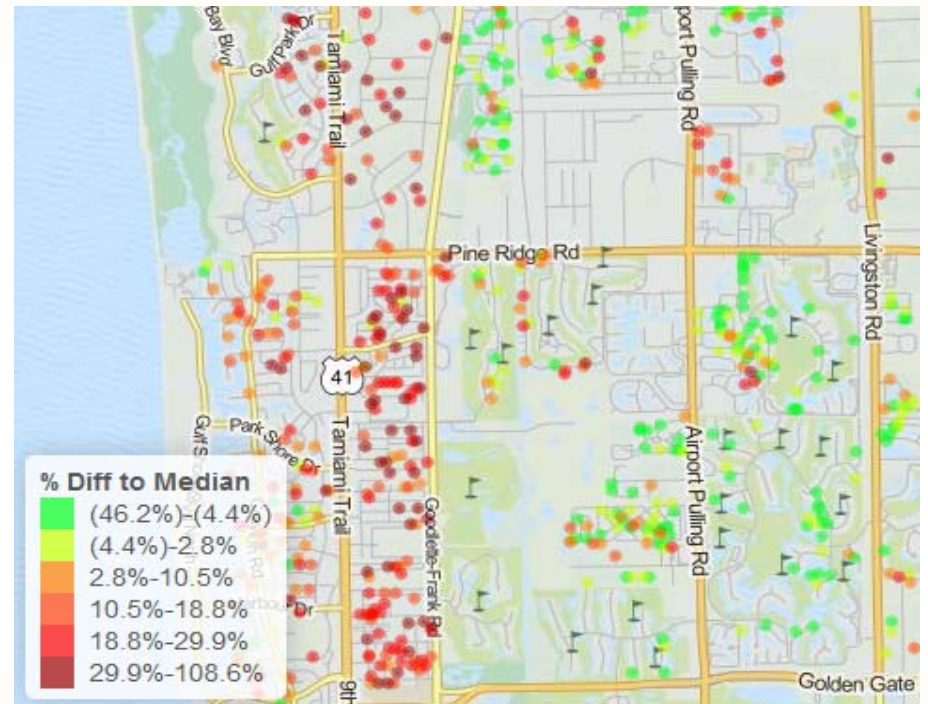- Actuarial Data

- Geospatial Data

- Technology and Application

DATA VISUALIZATION BASICS

What is data visualization?

- **Graphical representation of data**
  - Tables, bar charts, scatter plots, etc.
- **Data visualization is both art and science**

Lon	Lat	Diff
30.40686	-84.2982	1.052
30.40666	-84.2971	0.688
30.40686	-84.2975	0.331
30.40634	-84.2956	0.812
:	:	:

Why visualize data?

- So that **you** can understand it better
  - Efficiency and new perspectives
  - Useful for every step of the analysis (exploration, QC, validation)

- To communicate more effectively with **someone else**
  - Note: communication isn't about you!
  - Focus on your audience:
    - What's in it for them?
    - What do you want them to "take away"?
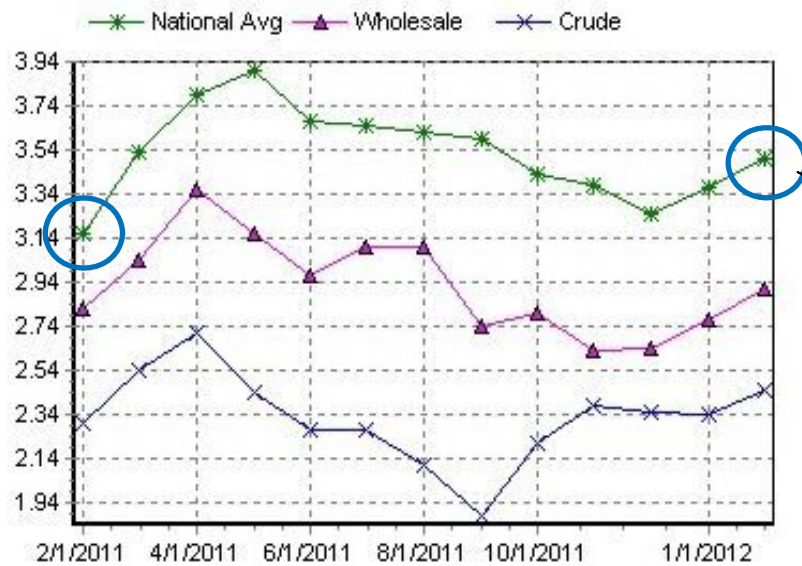   - What context will help them "get it"?

Make it easy to understand your point!

- Show **comparisons**, contrasts, differences
  - Encourage eye-to-eye comparison
  - Use "small multiples"
  - Highlight key points
- Show mechanism, explanation, **systematic structure**
  - Have a clear purpose
  - Show multiple levels of detail
- **Integrate** words, numbers, images, diagrams
  - Callouts can emphasize key take-aways
- Describe the **evidence**: titles, scales, sources, issues
- **Content (= meaning)** counts most of all

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

Basic principles

- **Honesty**
- Remove junk / ink
- Revise, revise, revise

Don't distort your data!
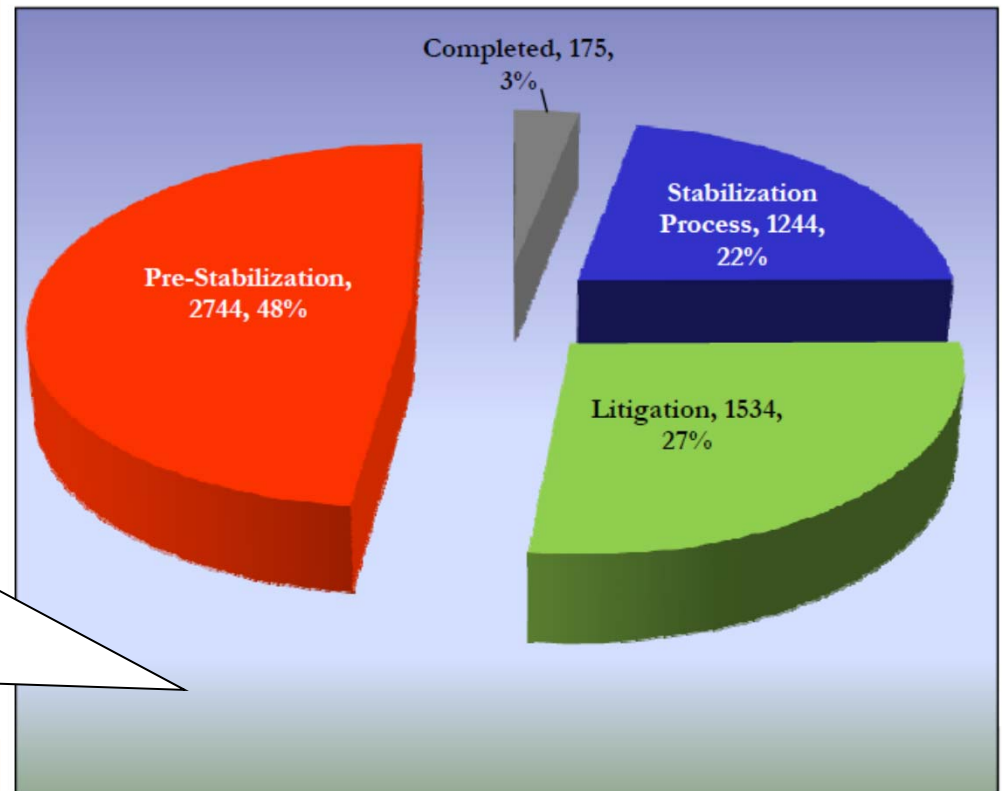
Basic principles

- Honesty
- **Remove junk / ink**
- Revise, revise, revise

Pie charts = hard to read
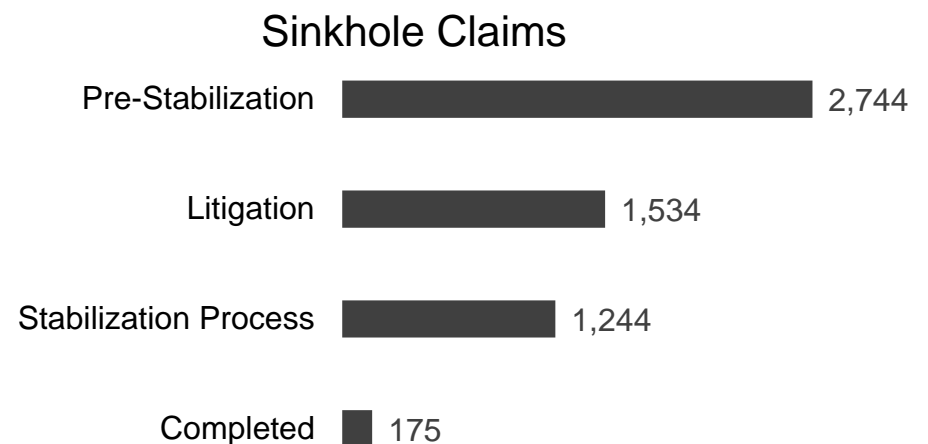- 3D + blue gradient background = even harder
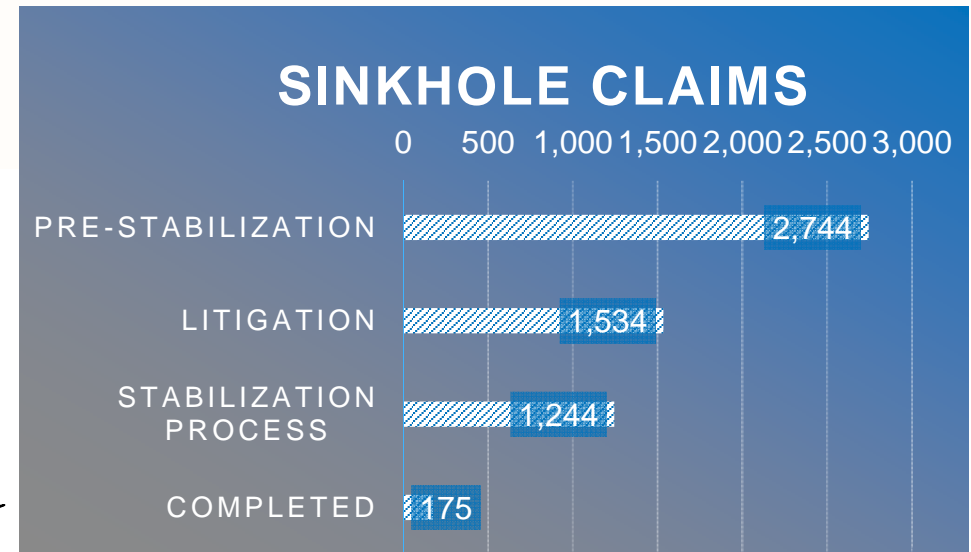
Really only 4 numbers here
- Bar chart or table would be more effective

Completed, 175, 3%

Stabilization Process, 1244, 22%

Pre-Stabilization, 2744, 48%

Litigation, 1534, 27%

Basic principles

- Honesty
- Remove junk / ink
- **Revise, revise, revise**

Don't settle for defaults! Sometimes creating a bare-bones graphic is more work, but the elegance and clarity is worth it.

SINKHOLE CLAIMS

	0	500	1,000	1,500	2,000	2,500	3,000
PRE-STABILIZATION						2,744	
LITIGATION			1,534				
STABILIZATION PROCESS			1,244				
COMPLETED	175						

Sinkhole Claims

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

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
- Pie charts
- Let type oppress graphics
- No ALL CAPS, **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'ts of Presenting Data, Facts, and Figures*

ACTUARIAL DATA

Data set 1: premium volume

LOB	2014 Prem	2015 Prem	Growth
1	6,682,095	8,372,665	25.3%
2	709,789	848,198	19.5%
3	18,998,022	19,036,018	0.2%
4	6,134,878	6,300,520	2.7%
5	729,498	938,135	28.6%
6	108,983	139,062	27.6%
7	1,317,246	1,625,482	23.4%
8	4,896,417	5,616,190	14.7%
9	1,385,081	1,750,743	26.4%
10	731,233	930,859	27.3%
Total	41,693,243	45,557,872	10.4%

Data set 1 (easier to read)

LOB	Premium ($M)		Growth
---	2014	2015	---
1	6.7	8.4	25.3%
2	0.7	0.8	19.5%
3	19.0	19.0	0.2%
4	6.1	6.3	2.7%
5	0.7	0.9	28.6%
6	0.1	0.1	27.6%
7	1.3	1.6	23.4%
8	4.9	5.6	14.7%
9	1.4	1.8	26.4%
10	0.7	0.9	27.3%
Total	**41.7**	**45.6**	**10.4%**

Goal: compare portfolio composition

2014 Premium Split
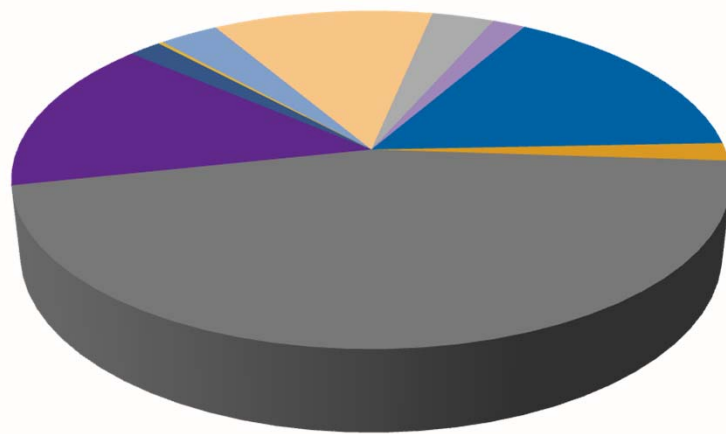
2015 Premium Split
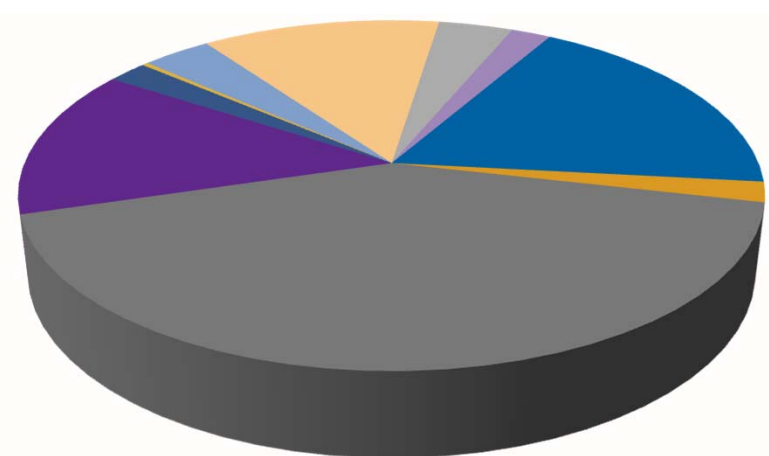
Legend:
- LOB 1
- LOB 2
- LOB 3
- LOB 4
- LOB 5
- LOB 6
- LOB 7
- LOB 8
- LOB 9
- LOB 10

- Unnecessary ink (3rd dimension is meaningless)
- Distortion creates hurdle to understanding

Goal: compare portfolio composition
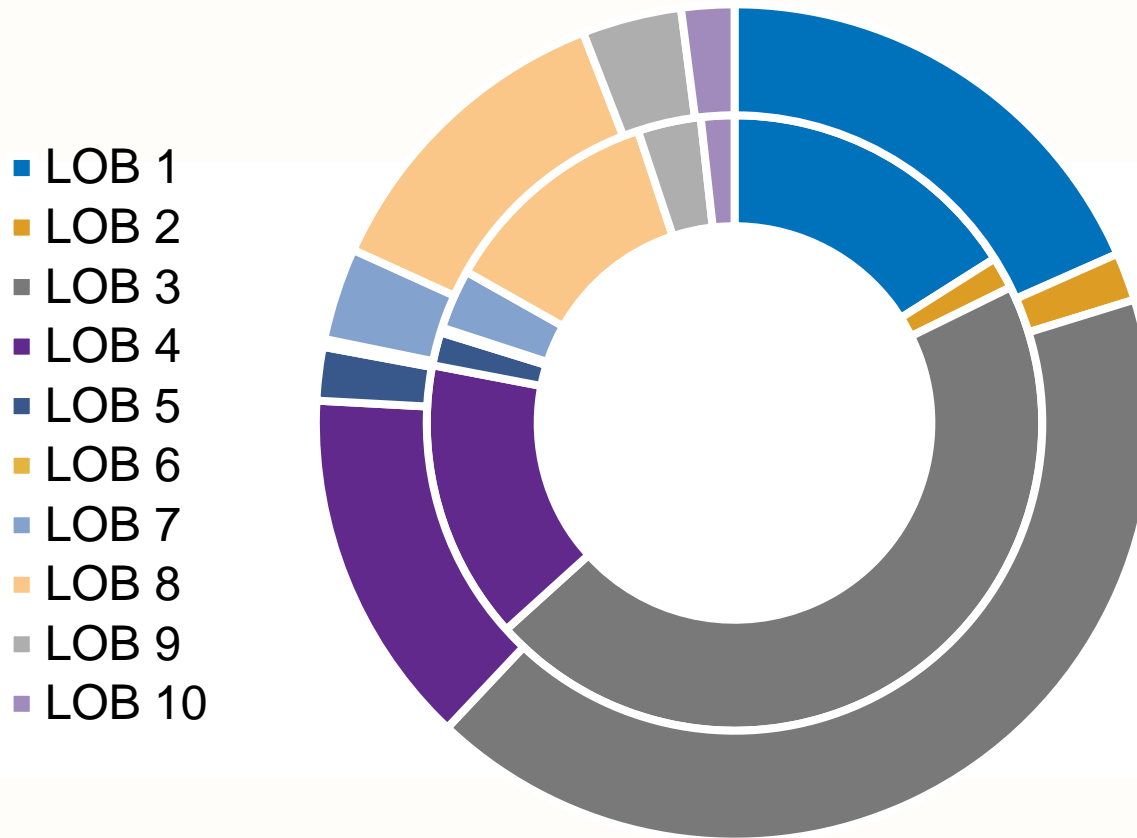
2014 Premium Split
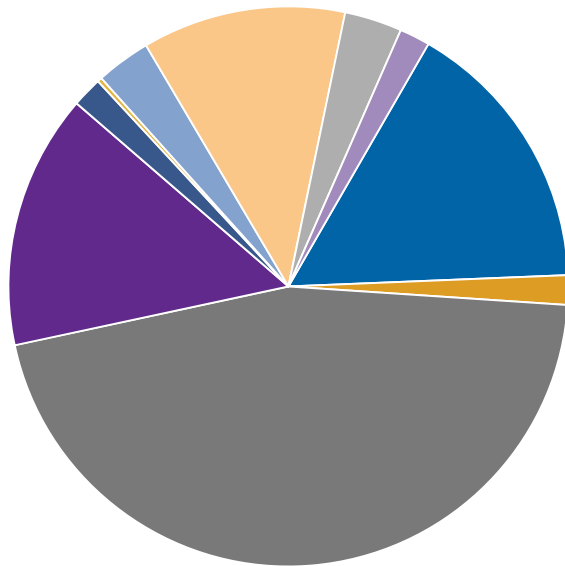
2015 Premium Split
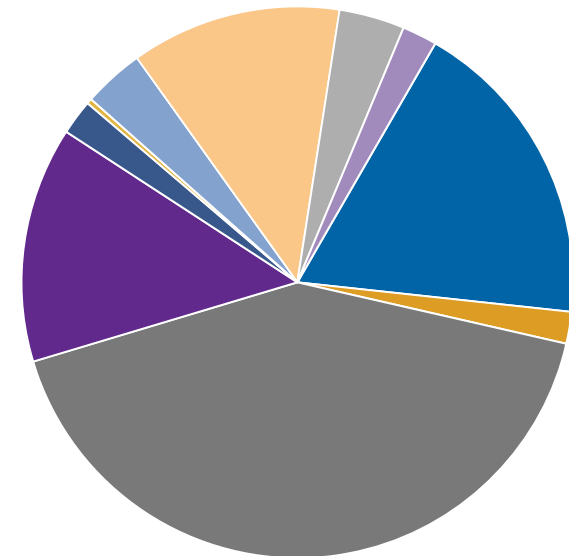
Legend:
- LOB 1
- LOB 2
- LOB 3
- LOB 4
- LOB 5
- LOB 6
- LOB 7
- LOB 8
- LOB 9
- LOB 10

- Easier to compare, but still not easy!
- How to make it easier?

Goal: compare portfolio composition

- LOB 1
- LOB 2
- LOB 3
- LOB 4
- LOB 5
- LOB 6
- LOB 7
- LOB 8
- LOB 9
- LOB 10

- Did LOB 4 increase or decrease as a share of the overall premium?

Goal: compare portfolio composition

Legend:
- LOB 1
- LOB 2
- LOB 3
- LOB 4
- LOB 5
- LOB 6
- LOB 7
- LOB 8
- LOB 9
- LOB 10

Chart values:
- 18%
- 16%
- 12%
- 12%
- 15%
- 14%
- 45%
- 42%
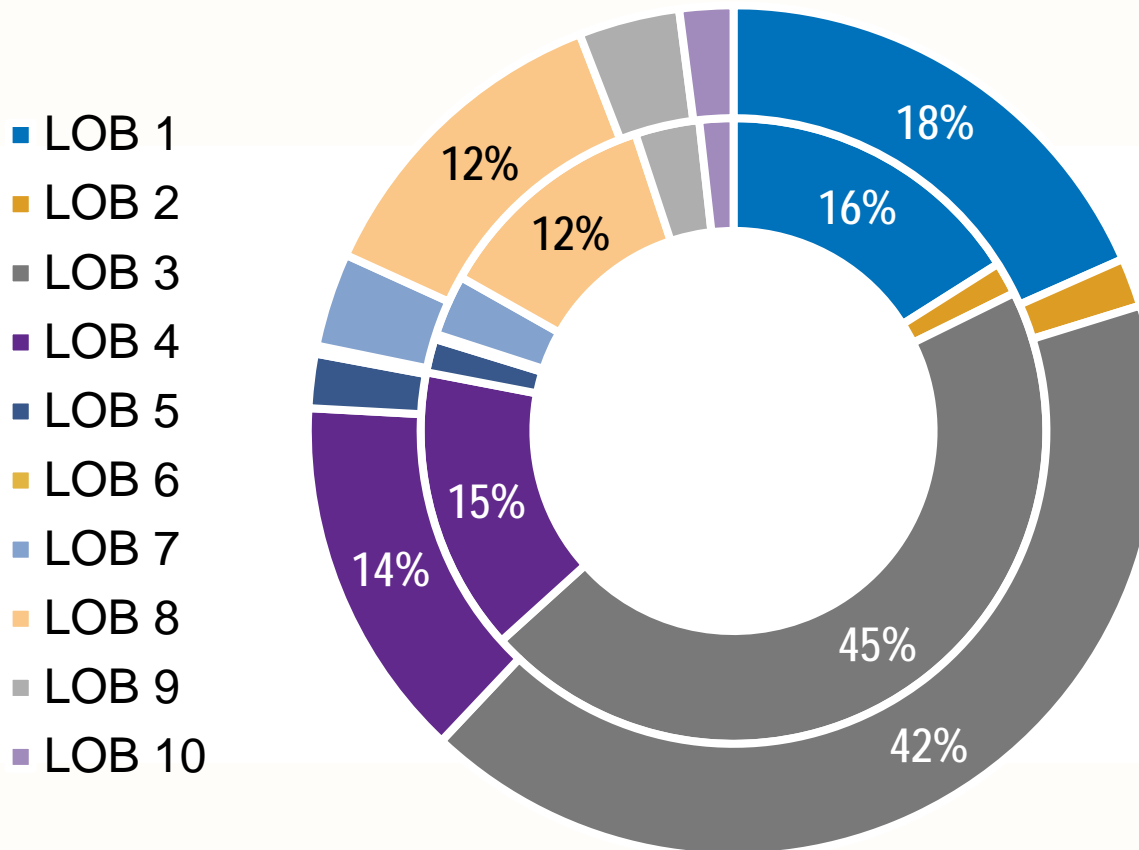
- Radial angles are difficult to judge
- Relative areas are misleading

Goal: compare portfolio composition
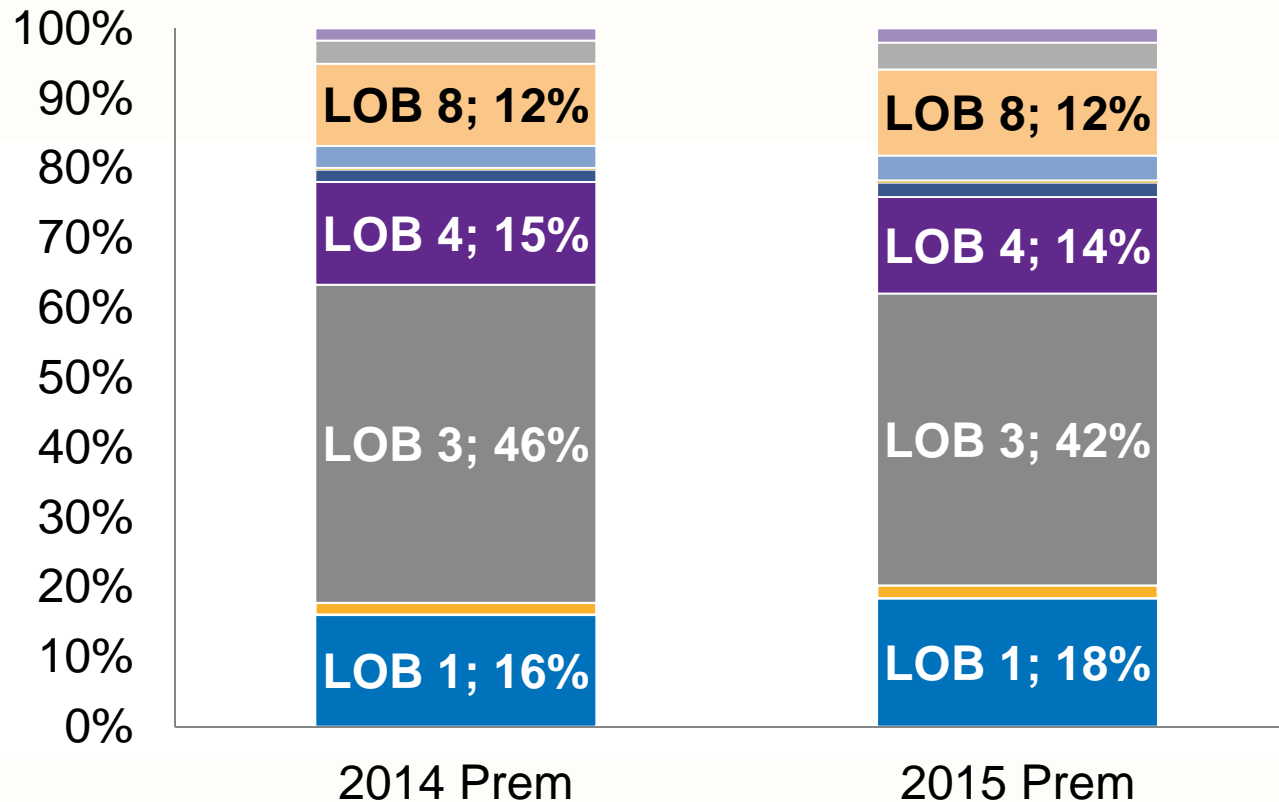
- No need to judge angles
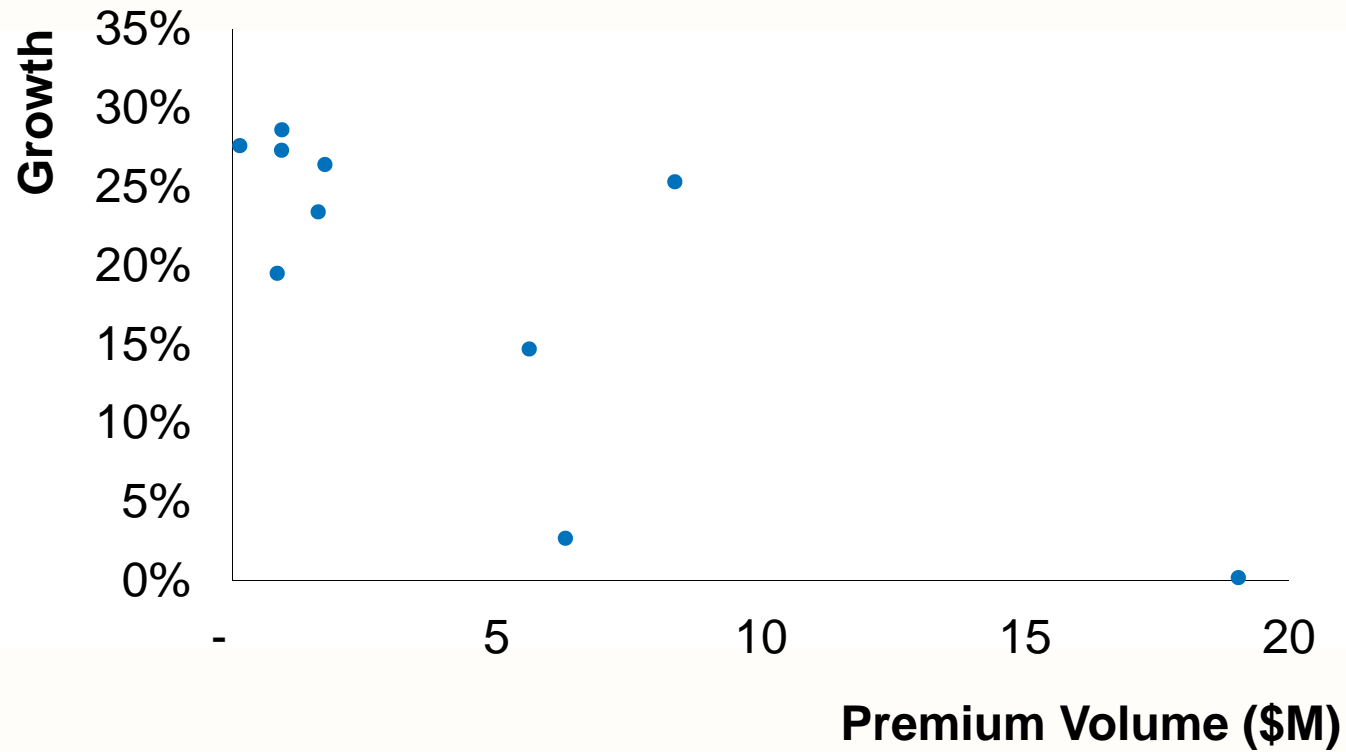- Areas no longer mislead
- Labels make it even easier

New goal: understand growth by LOB
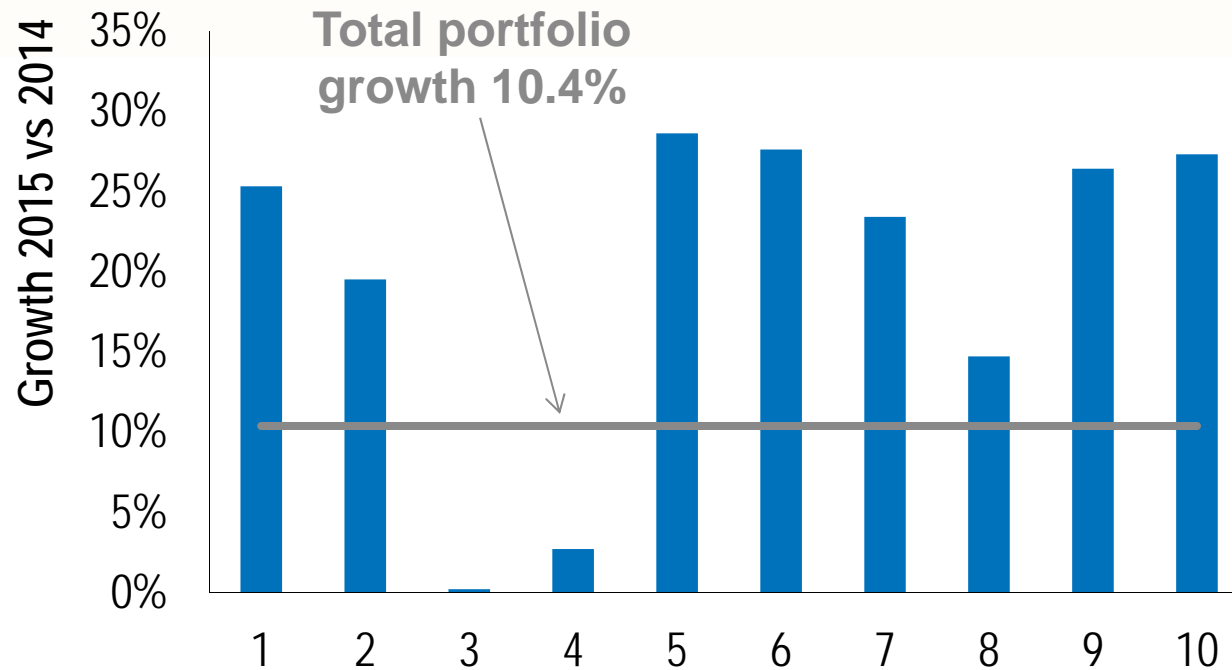
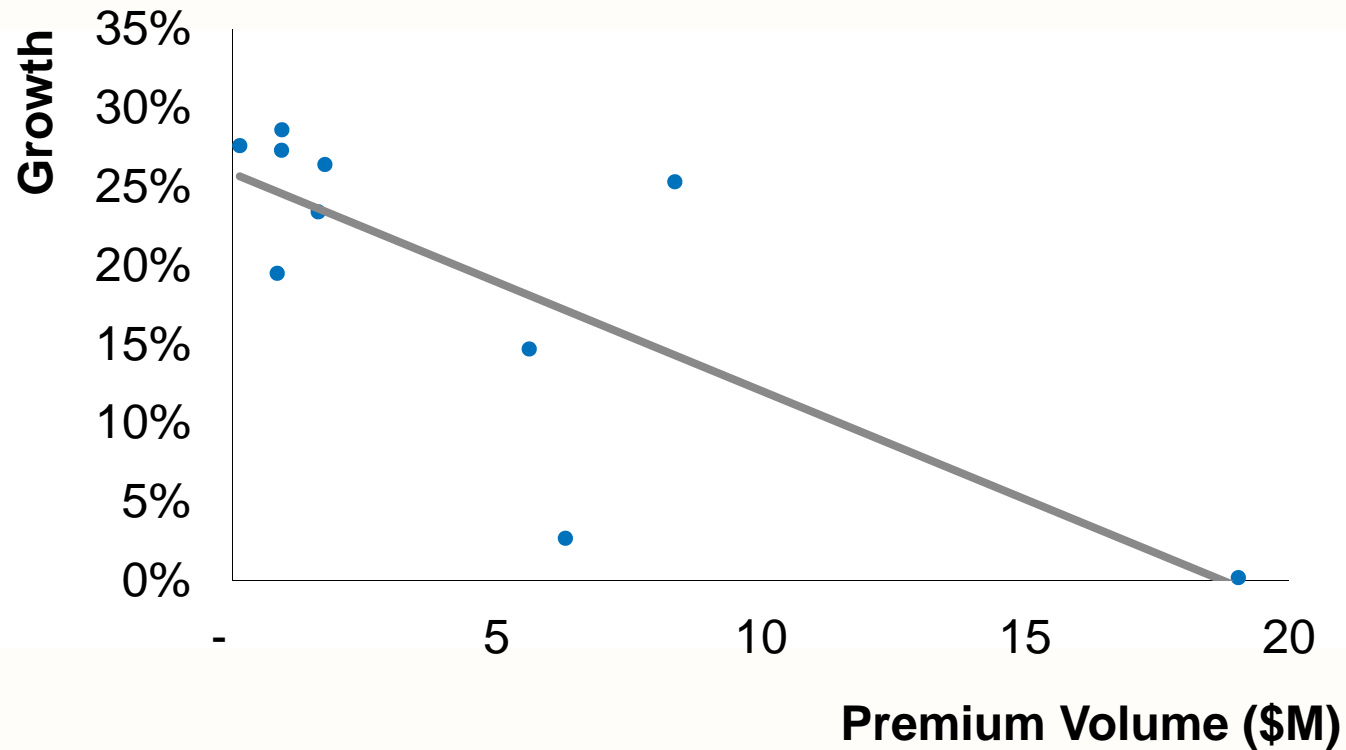- Focused on the point: growth
- Line and callout aid comparison

New goal: understand growth by LOB

- Additional dimension adds meaning

23

New goal: understand growth by LOB

- Additional dimension adds meaning
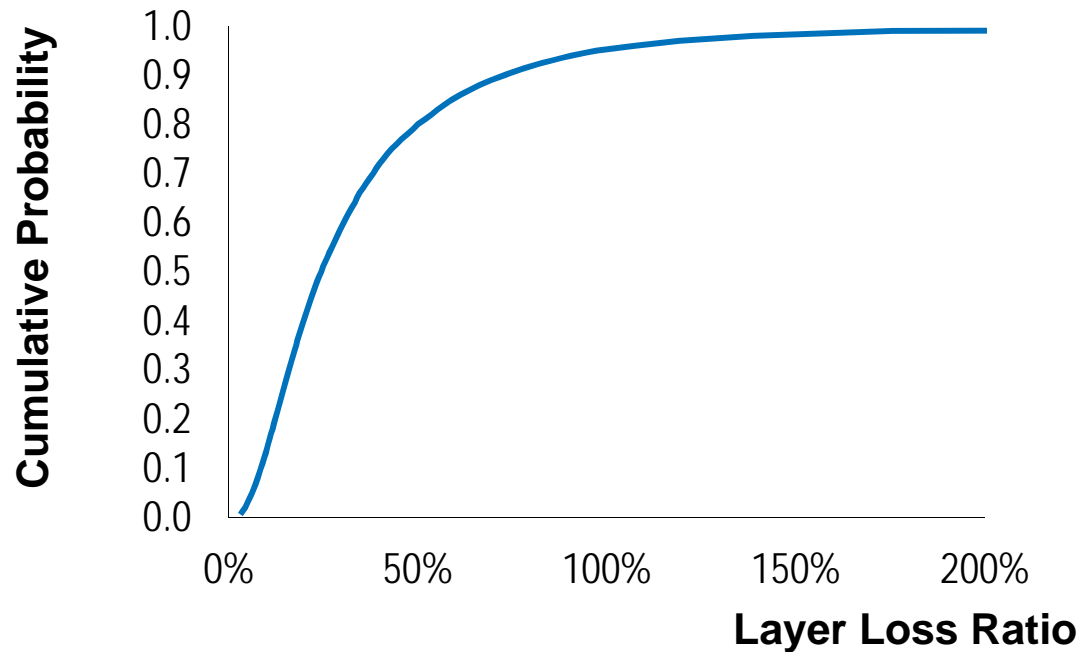- Trend line may shed further light

24

Data set 2: simulation results

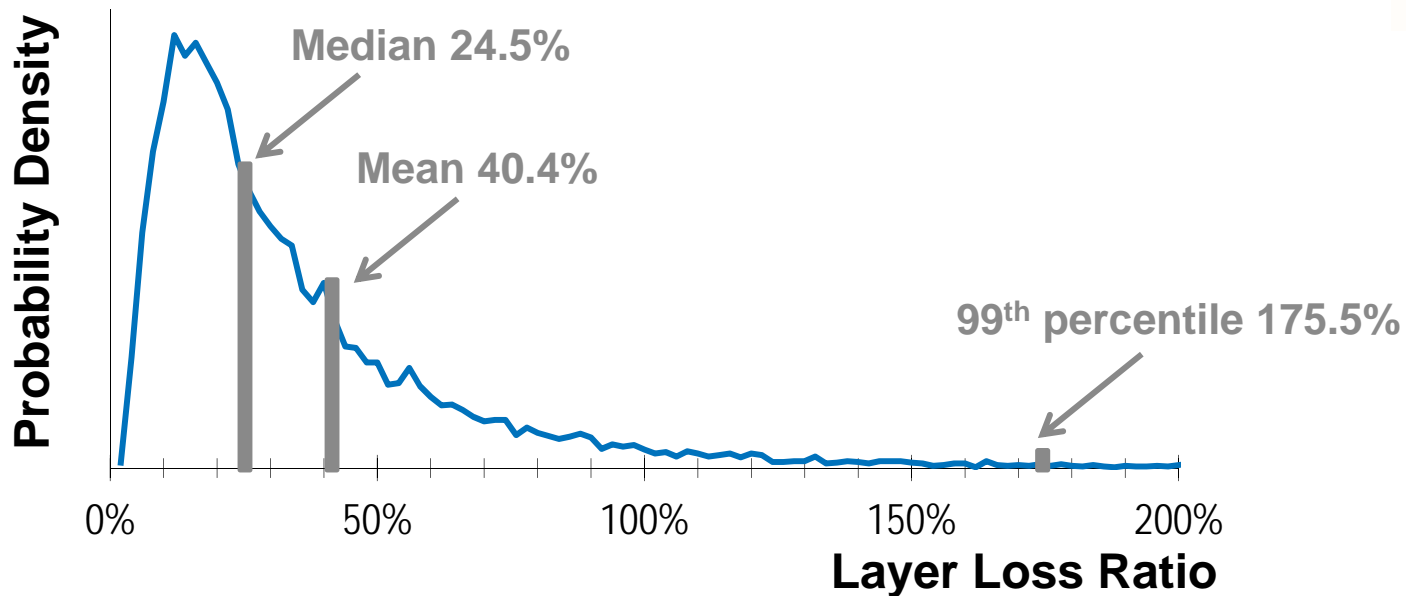| Simulation #	Simulated Layer Loss Ratio
1	0.9%
2	1.6%
3	1.6%
4	1.7%
5	1.8%
⋮	⋮
9996	391.2%
9997	397.3%
9998	407.0%
9999	441.2%
10,000	683.3%

How to visualize?

- Why? It's easy for the spreadsheet jockey to make
- But it's definitely <u>not</u> easy for the non-technical audience to read
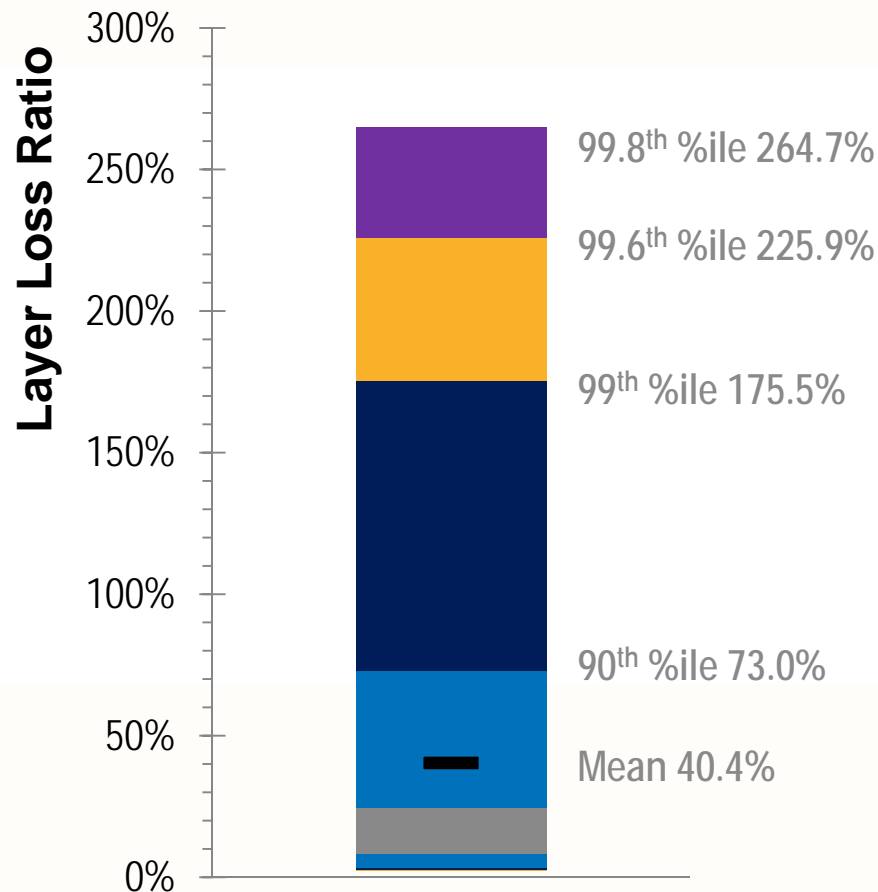
Goal: understand overall distribution

- **May better convey a sense of distribution behavior**
- **But…**
  – Difficult for non-technical audience to understand Y axis
  – Does not help understand tail of distribution

Goal: understand overall distribution

Layer Loss Ratio

- 300%
- 250% — 99.8th %ile 264.7%
- — 99.6th %ile 225.9%
- 200%
- — 99th %ile 175.5%
- 150%
- 100%
- — 90th %ile 73.0%
- 50% — Mean 40.4%
- 0%

- Tukey bars
  - Can be easier for non-technical audience to interpret
  - May better illustrate "fatness" of the tail
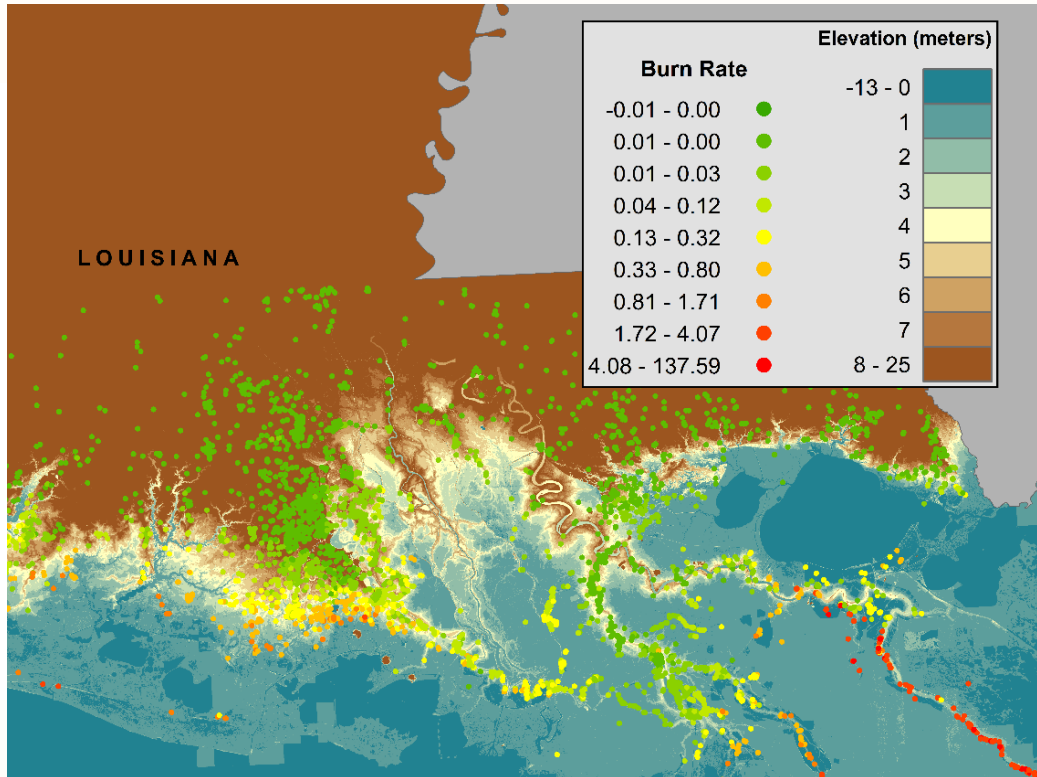- Providing the "tabular" data as callouts still helpful

GEOSPATIAL DATA

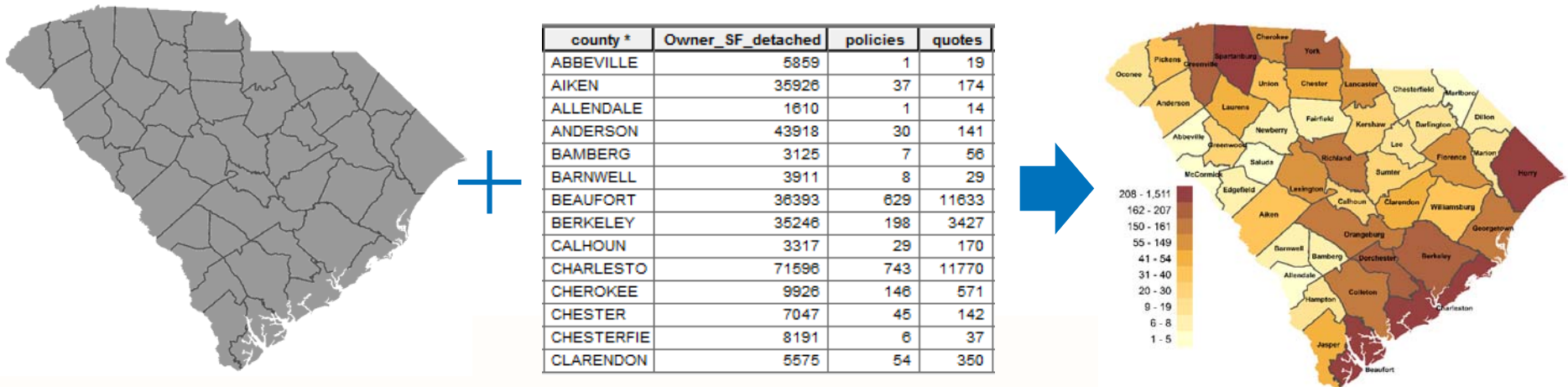# Why maps?

Elevation and burn rate by policy location

Burn rate = expected AAL per $1,000 of HO Coverage A

- Risk and exposure vary with geography

- Quick analysis of large volumes of data

- Visualization and analysis of spatial relationships that cannot be seen with tables and charts

GIS basics

- Geographic Information Systems (GIS)
  - Used to capture, store, edit, analyze, and present geographic data
- Map components
  - Software + spatial data (base map and/or areas to map) + data to visualize

county *	Owner_SF_detached	policies	quotes
ABBEVILLE	5859	1	19
AIKEN	35926	37	174
ALLENDALE	1610	1	14
ANDERSON	43918	30	141
BAMBERG	3125	7	56
BARNWELL	3911	8	29
BEAUFORT	36393	629	11633
BERKELEY	35246	198	3427
CALHOUN	3317	29	170
CHARLESTO	71596	743	11770
CHEROKEE	9926	146	571
CHESTER	7047	45	142
CHESTERFIE	8191	6	37
CLARENDON	5575	54	350

- Example of GIS polygon dataset or "layer" and corresponding table
- Number of policies by county can be used to classify and color each county area

31

Spatial data sources

- **U.S. Census TIGER files**
  - Political boundaries and base map info
    https://www.census.gov/geo/maps-data/data/tiger-line.html

- **U.S. Geological Survey**
  - National Hydrology Dataset (NHD)
    http://nhd.usgs.gov/
  - National Elevation Dataset (NED)
    - http://nationalmap.gov/elevation.html
  - National Land Cover Database (NLCD)
    - http://www.mrlc.gov/

- **NOAA**
  - Coastlines
    https://www.ngdc.noaa.gov/mgg/shorelines/
  - National Weather Service
    http://www.nws.noaa.gov/gis
  - National Hurricane Center
    http://www.nhc.noaa.gov/gis/

- **State and local agencies**
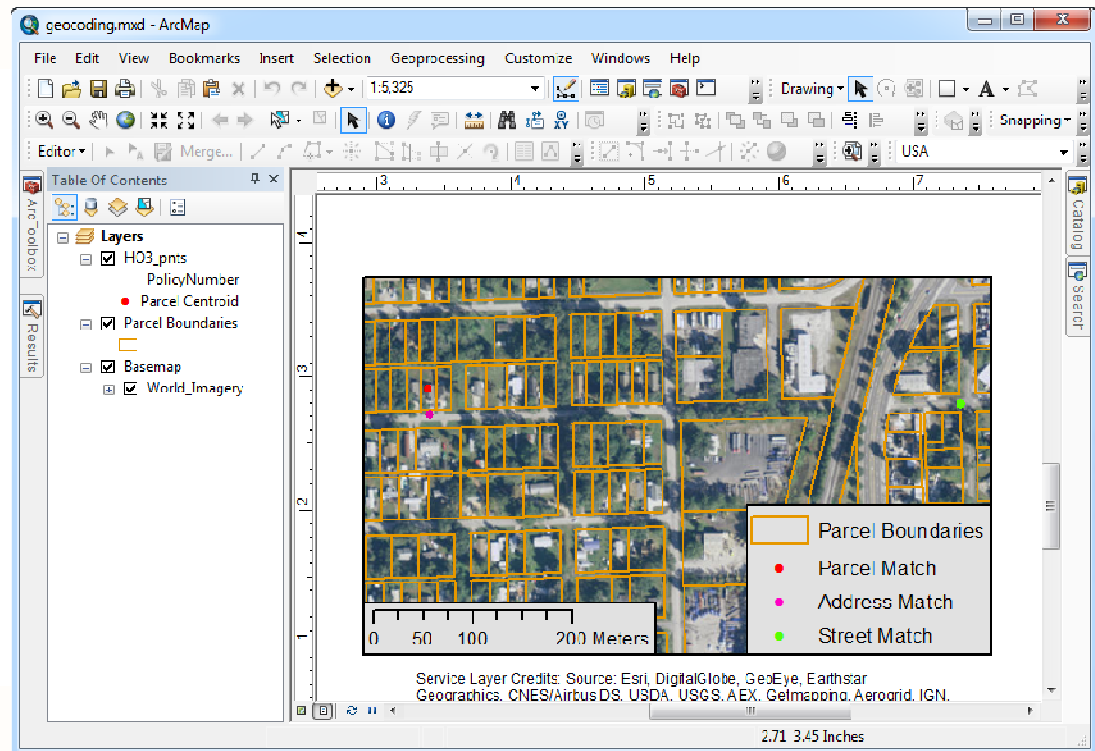 - E.g. parcels and land records from county assessors
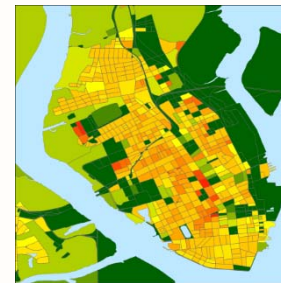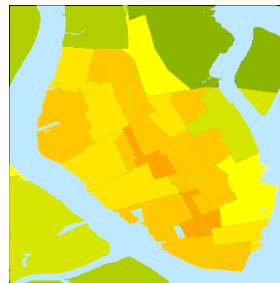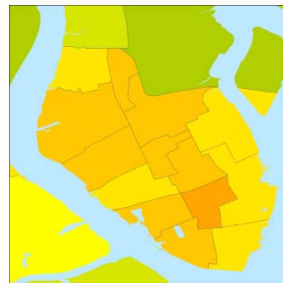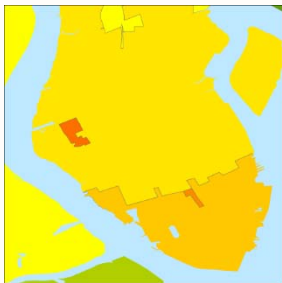
Population density in Charleston, SC plotted using ZIP Code Tabulation Areas (ZCTAs), Census Tracts, Block Groups, and Blocks.

Mapping tools: desktop GIS

- **ArcGIS for Desktop**
  - Part of ESRI's ArcGIS Platform
 - Licensing varies ($3K – $14K)

- **GRASS/QGIS**
  - Open Source

Desktop GIS applications like ArcGIS are useful for designing sophisticated map layouts and designs

Mapping tools: software with GIS

- **SAS/GIS and SAS/GRAPH**
  - Additional license
- **R**
  - Open Source
  - Lots of mapping libraries: ggplot2, maptools, ggmap…

    https://cran.r-project.org/web/views/Spatial.html
- **ArcGIS for Office**
  - ESRI Plug-In for MS Office
  - Included with ArcGIS Online
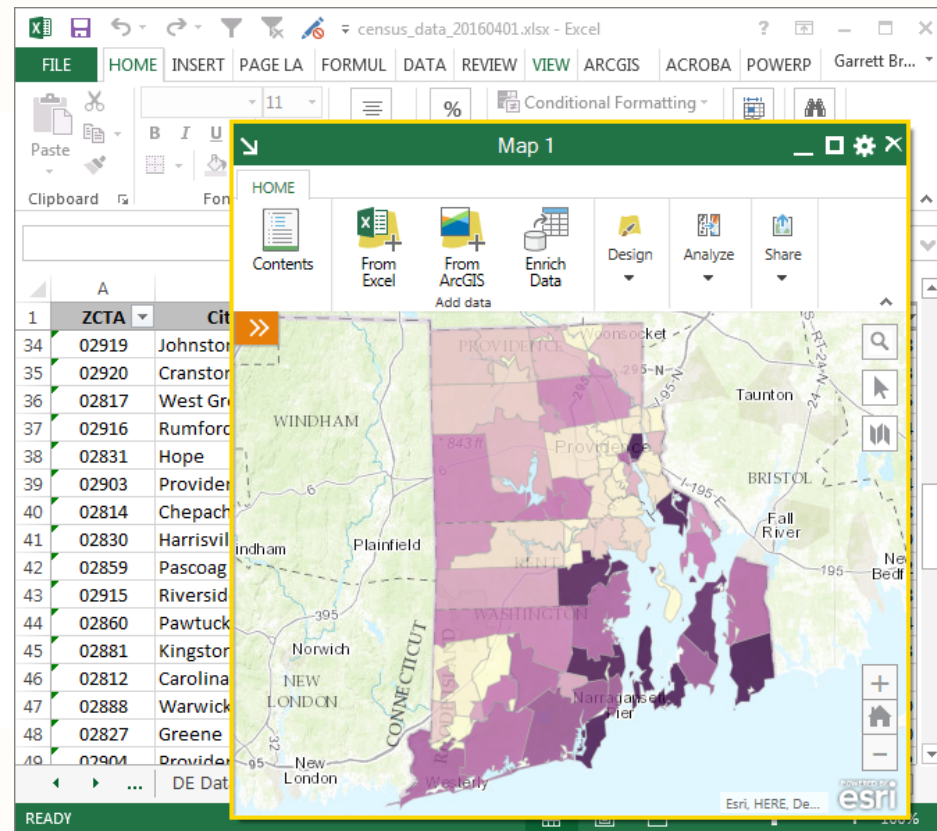
Excel with ArcGIS for Office Plug-In being used to visualize high-value ZCTAs in Rhode Island

Mapping tools: web maps

- **CartoDB**
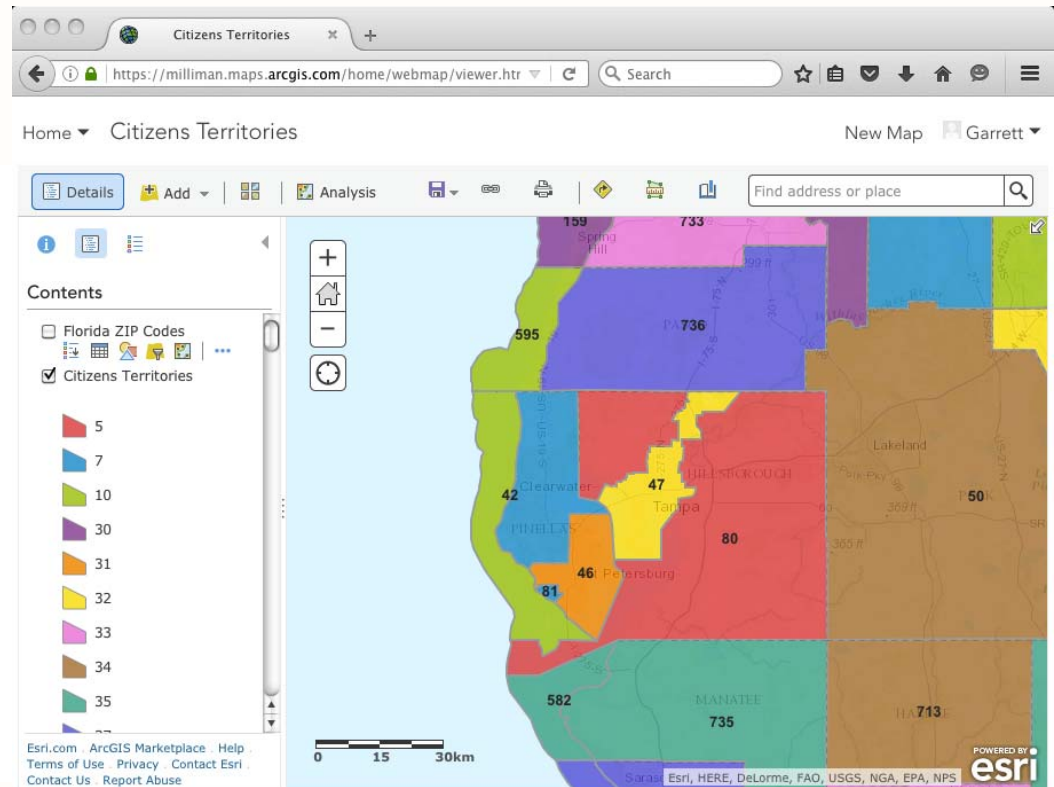 - Online GIS tool for visualization and analysis
- **Leaflet**
  - Open Source JS library for webmaps
  - Can be used with spatial libraries in R to create interactive maps
- **Google Maps**
- **SpatialKey**
- **ArcGIS Online**
  - Part of ESRI Platform
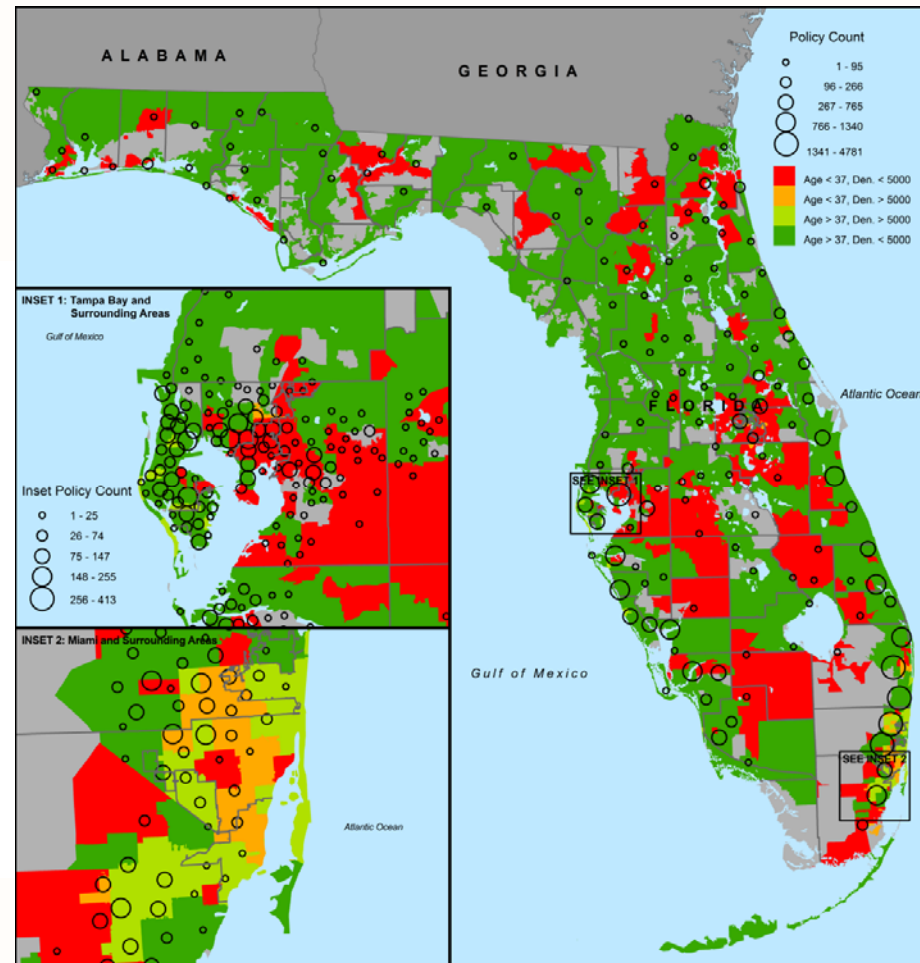 - Included with Server or ArcGIS for Desktop

ArcGIS Online being used to visualize custom rating territories

Map types

- Reference maps
  - Road maps, atlases, globes…
  - Avoid using reference maps as a background where possible
- Thematic maps
  - Choropleth (shaded-area)
  - Proportional symbol
  - Isarithmic or isopleth (contour map or heat maps)
  - Dot or dot density
  - Dasymetric

Combination of proportional symbols and shaded area allows two different fields to be shown on the same map.

This map shows results of a segmentation analysis on profitability (red = unprofitable, green = profitable) along with policy count as a graduated circle.

36

Use of color

- **Great guide for color selections:**
  - http://colorbrewer2.org/

- **Use color effectively**
  - Single hue is best for sequential data
  - Avoid political or biased color palettes if not related to message at hand
  - Avoid red-green color ramps

Total Policies In-Force by County South Carolina
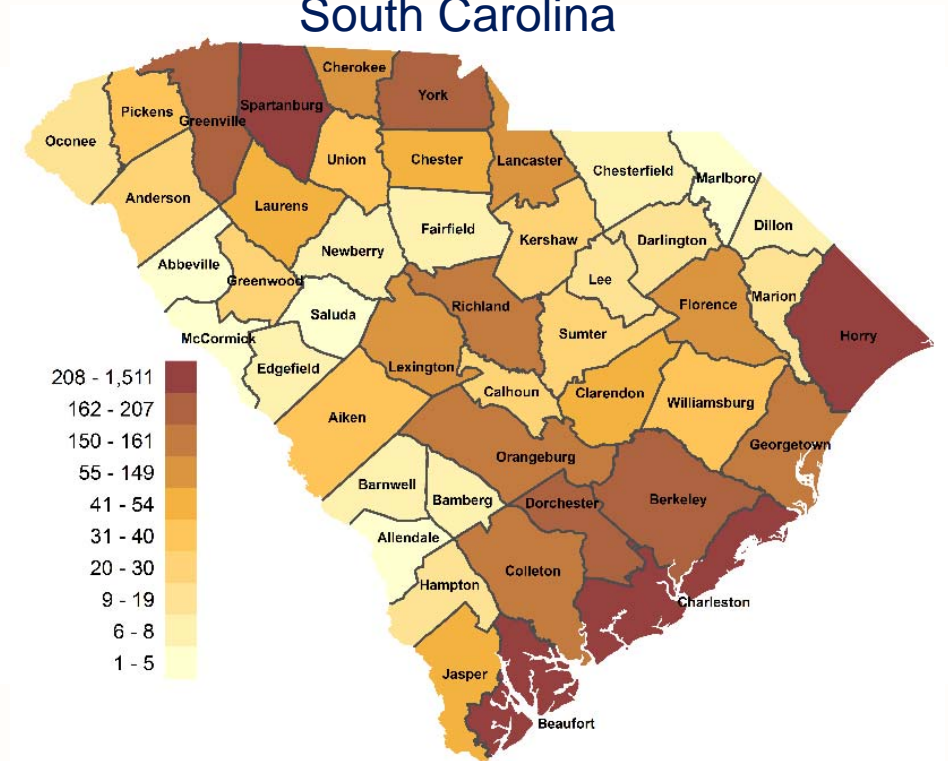
Mapping considerations

- **Map components**
  - Title, legend, scale, graticule, neatline, symbols, labels…
  - Most thematic maps do not need all these components

- **Generalization**
  - Selection, simplification, classification, symbolization

- **Balance and hierarchy**
  - Create clear figure-ground relationship
  - Balance all items a page
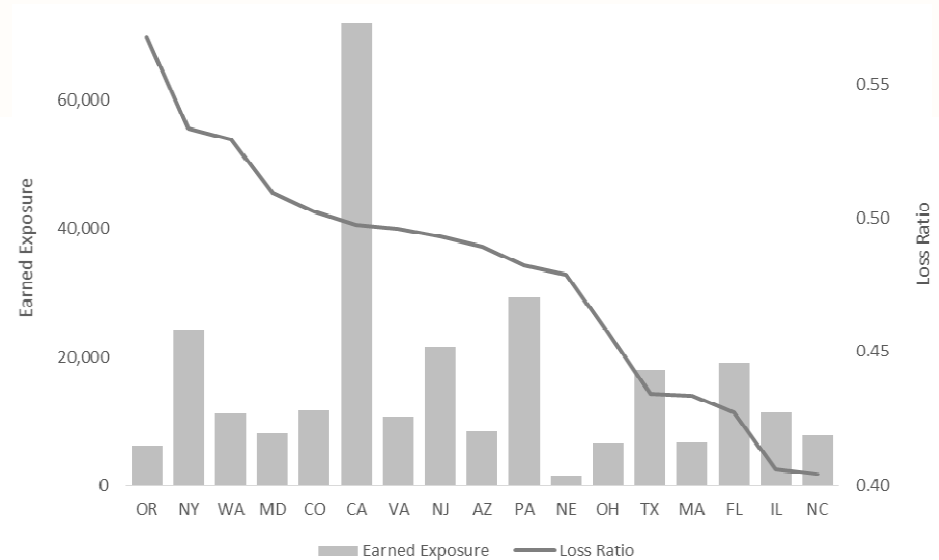 - Create clear hierarchy in symbolization and labeling
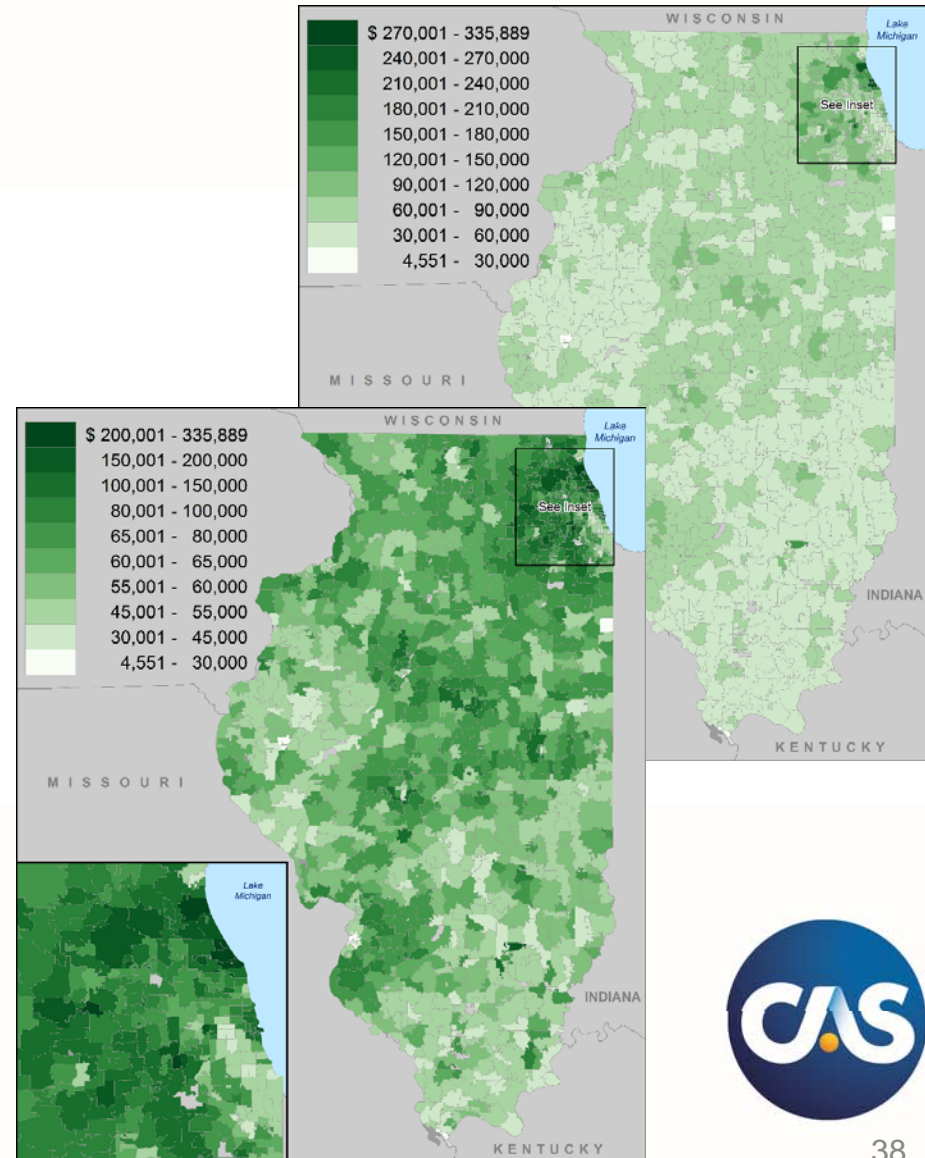
TECHNOLOGY AND APPLICATION

Finding the right tool

- **Desktop Applications**
  - Excel, ArcMap, Illustrator
- **Coded solutions**
  - Matlab, R, SAS, Python
- **Web Applications**
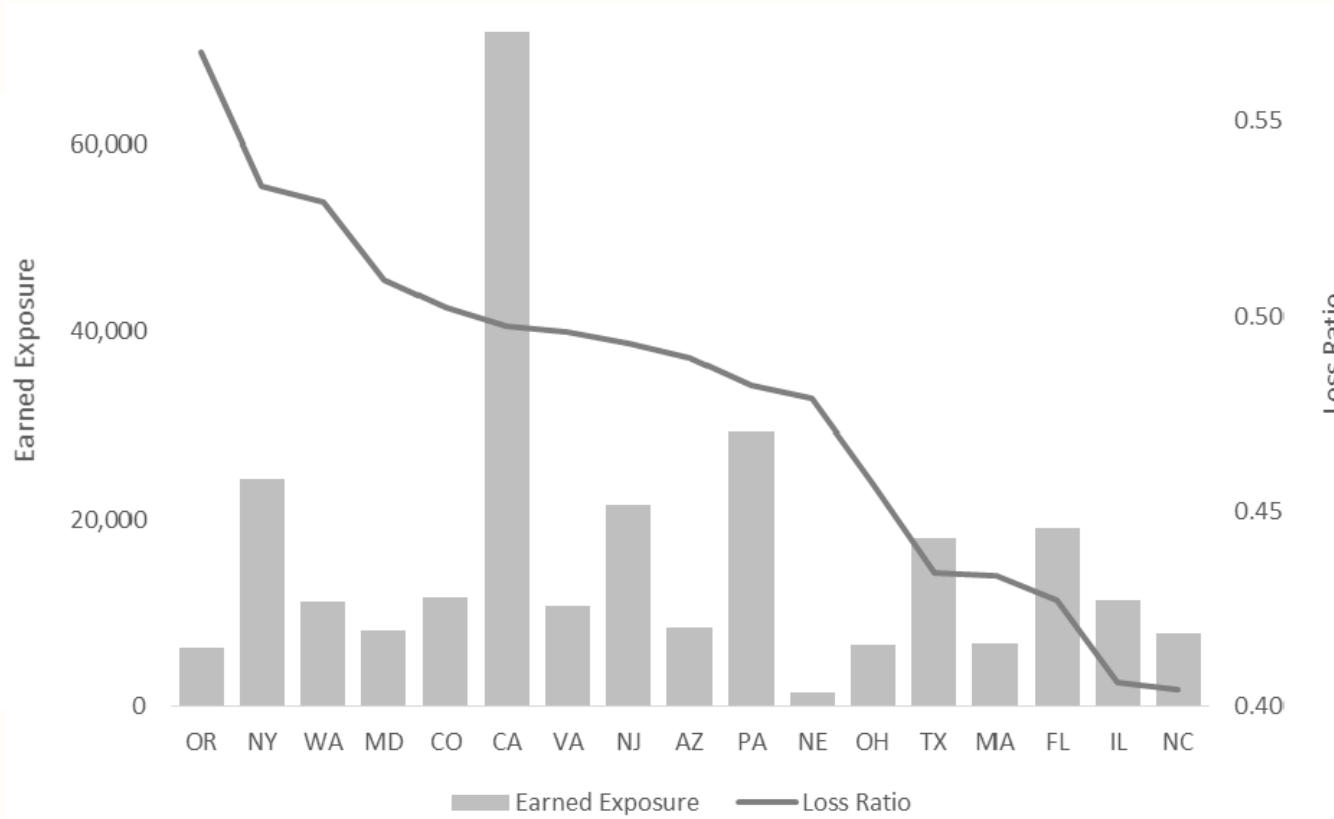 - Tableau, Google Chart Tools, CartoDB, SpatialKey


```
proc sgplot data=input;
where pettype='Dog';
vbar state/response=EarnedExposure;
vline  state/response=lr y2axis;
run;
```
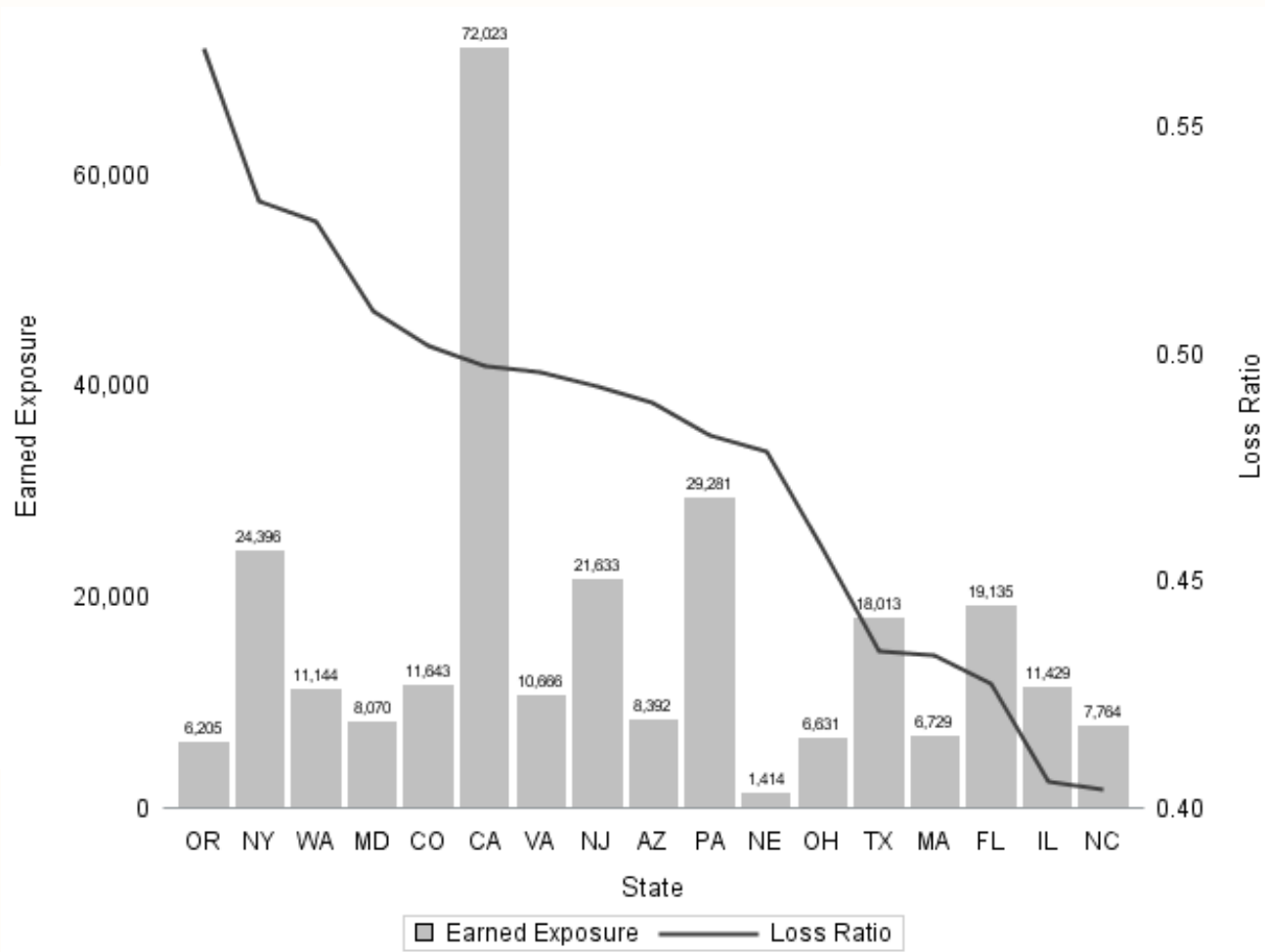
Finding the right tool: Excel, SAS, Google Charts

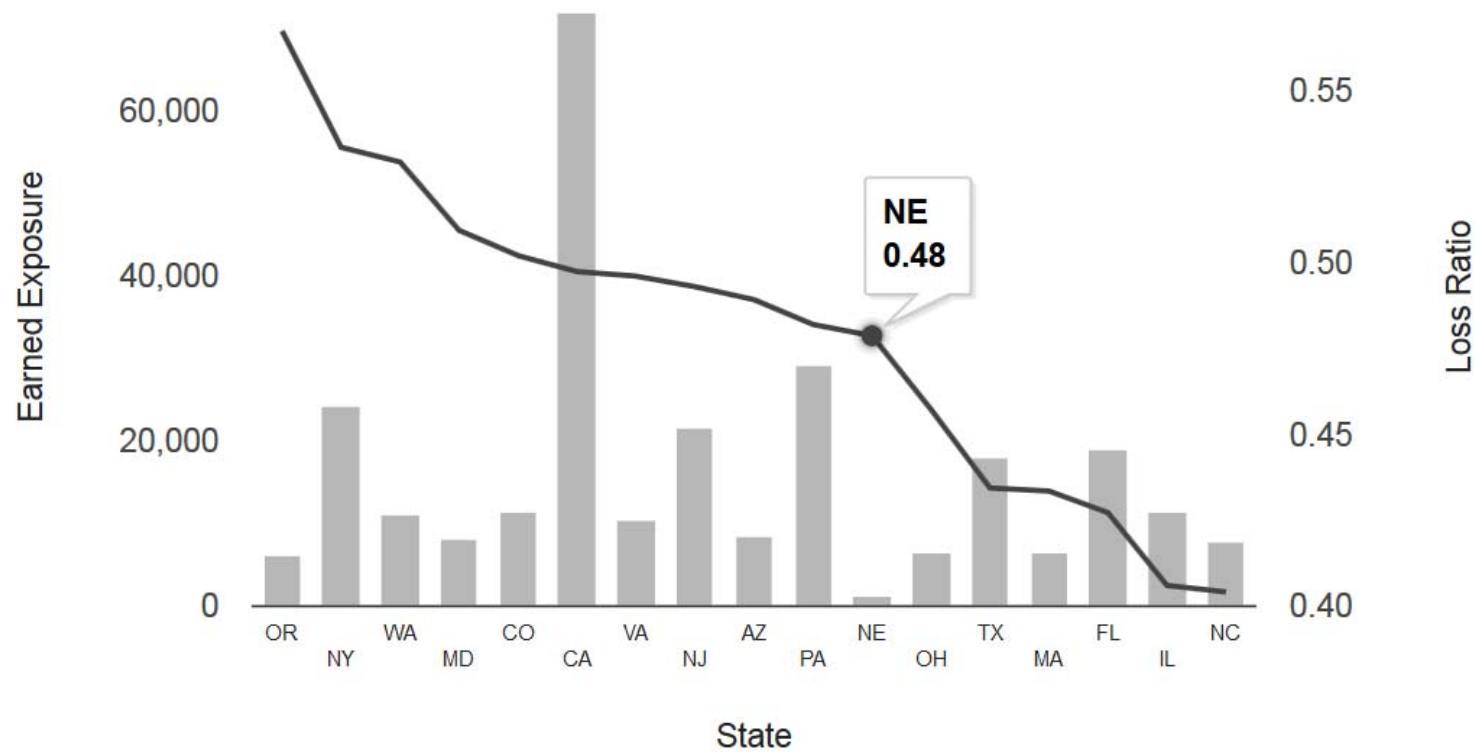# Finding the right tool: Excel

Finding the right tool: SAS

Finding the right tool: SAS

```
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;
```

44

Finding the right tool: Google Charts
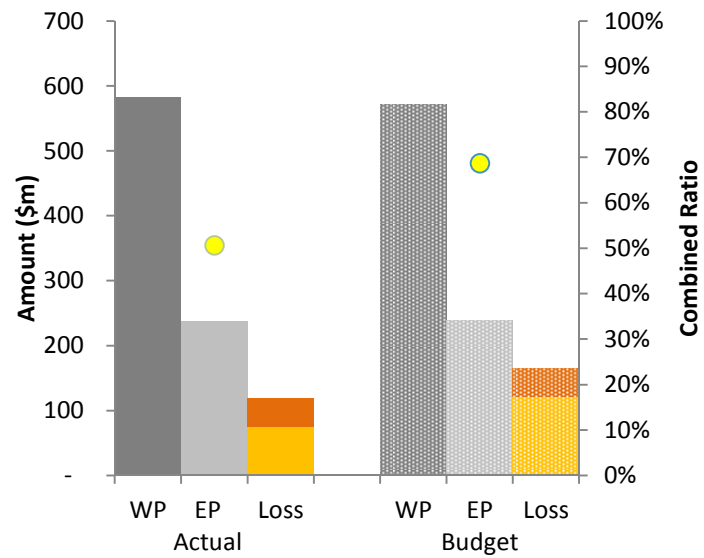

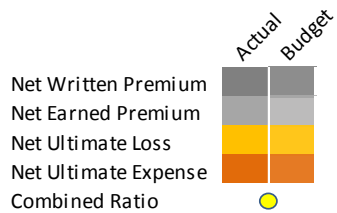Link to online graphic.

Finding the right tool: Excel, Tableau, R

Finding the right tool: Excel

2016 Q1 Actual versus Budget

Net UY 2015	WP	Actual EP	Loss	WP	Budget EP	Loss
Net Written Premium (Budget)				571		
Net Written Premium (Actual)	583					
Net Earned Premium (Budget)					239	
Net Earned Premium (Actual)		237				
Net Ultimate Loss (Budget)						121
Net Ultimate Loss (Actual)			75			
Net Ultimate Expense (Budget)						43
Net Ultimate Expense (Actual)			45			
Net Combined Ratio (Budget)					0.69	
Net Combined Ratio (Actual)		0.51				

Legend:
- Net Written Premium (Actual / Budget)
- Net Earned Premium (Actual / Budget)
- Net Ultimate Loss (Actual / Budget)
- Net Ultimate Expense (Actual / Budget)
- Combined Ratio

Finding the right tool: Tableau
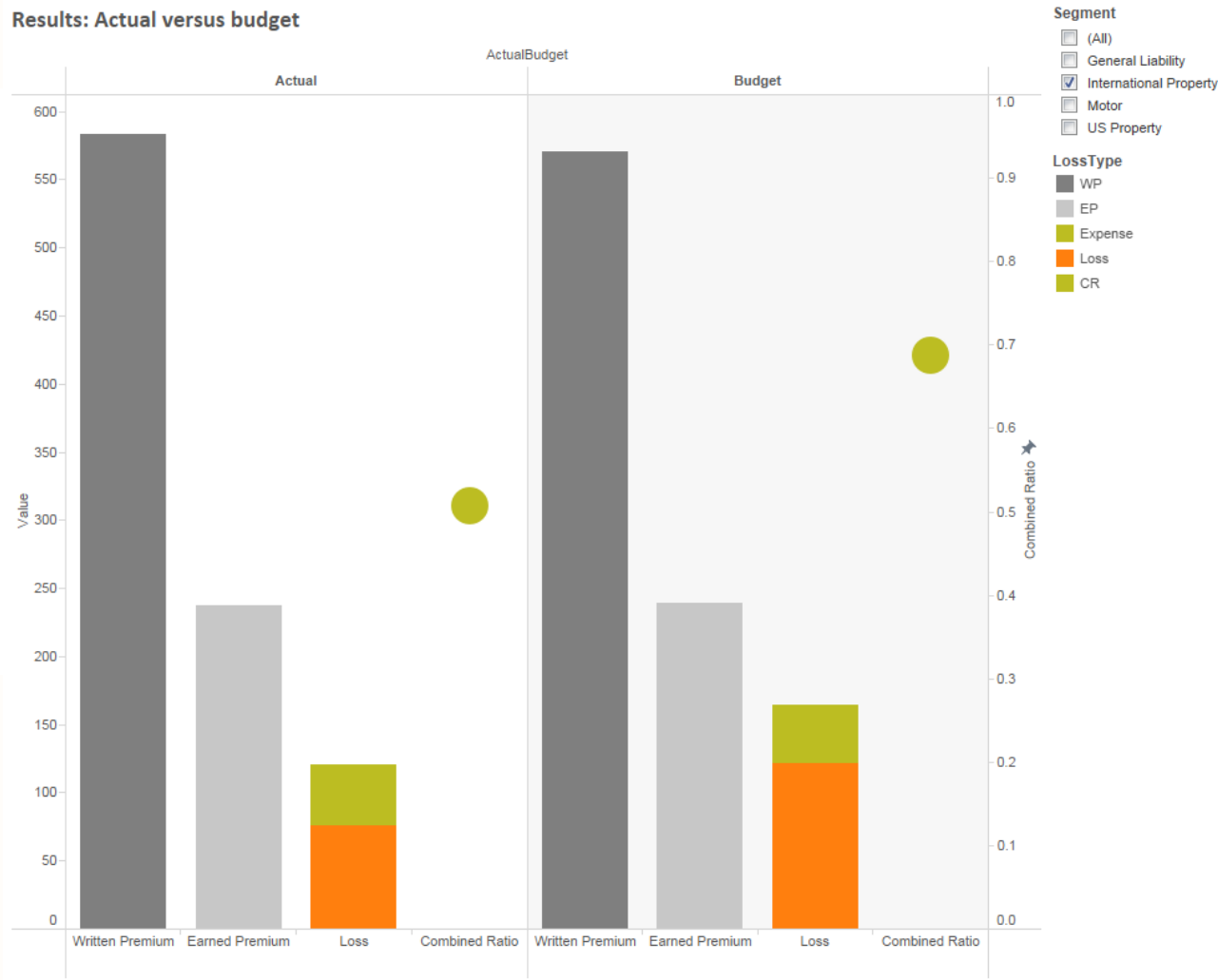
Finding the right tool: R

```
results <- read.csv("QTD for R mod.csv")
results <- results[results$ActualBudget == "Actual",]
p1 <- ggplot(results[results$Item!="(4)CR",], aes(x = Item, y = Value, fill = LossType)) +
  geom_bar(stat="identity", width = .6) +
  scale_fill_manual(values=c("#A9A9A9", "#DEDEDE", "#FFA500", "#FFFF00")) +
  labs(x="Item", y="Amount ($m)", title="Quarterly Results (Actual)") +
  theme_bw() +
  theme(legend.justification=c(0,1),
        legend.position=c(.6,1),
        plot.title=element_text(size=15,vjust=1),
        panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        panel.border = element_blank(),
        axis.text.x=element_text(size=10),
        axis.text.y=element_text(size=10),
        axis.title.x=element_text(size=10),
        axis.title.y=element_text(size=10))
p2 <- ggplot(results[results$Item=="(4)CR",], aes(x=Item, y = Combined.Ratio)) +
  geom_point(colour = "#FFFF00", size = 10) +
  ylim(0,1) +
  theme_bw() +
  labs(y="test") +
  theme(panel.border = element_blank()) %+replace%
  theme(panel.background = element_rect(fill = NA),
        panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        panel.border = element_blank(),
        axis.text.y=element_text(size=10),
        axis.title.y=element_text(size=10))

# extract gtable
g1 <- ggplot_gtable(ggplot_build(p1))
g2 <- ggplot_gtable(ggplot_build(p2))

# overlap the panel of 2nd plot on that of 1st plot
pp <- c(subset(g1$layout, name == "panel", se = t:r))
g <- gtable_add_grob(g1, g2$grobs[[which(g2$layout$name == "panel")]], pp$t,
             pp$l, pp$b, pp$l)

# axis tweaks
ia <- which(g2$layout$name == "axis-l")
ga <- g2$grobs[[ia]]
ax <- ga$children[[2]]
ax$widths <- rev(ax$widths)
ax$grobs <- rev(ax$grobs)
ax$grobs[[1]]$x <- ax$grobs[[1]]$x - unit(1, "npc") + unit(0.15, "cm")
g <- gtable_add_cols(g, g2$widths[g2$layout[ia, ]$l], length(g$widths) - 1)
Act <- gtable_add_grob(g, ax, pp$t, length(g$widths) - 1, pp$b)
```
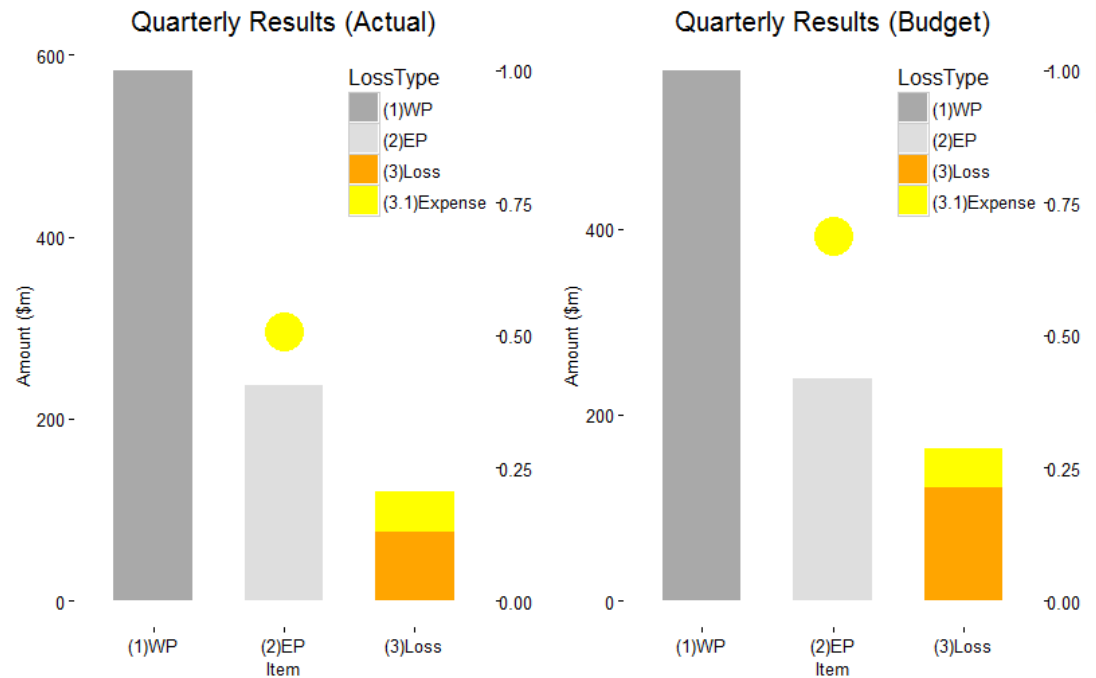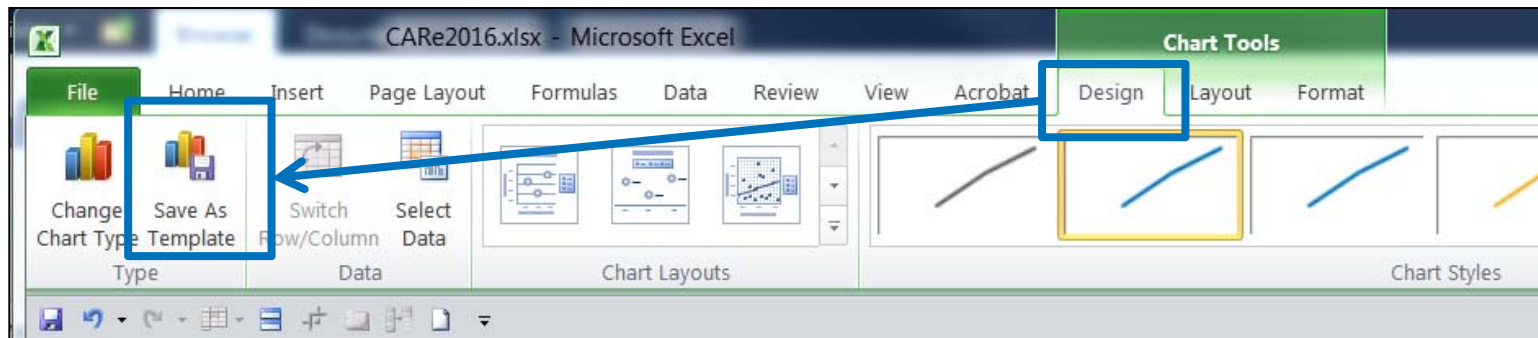
How to find Excel chart templates

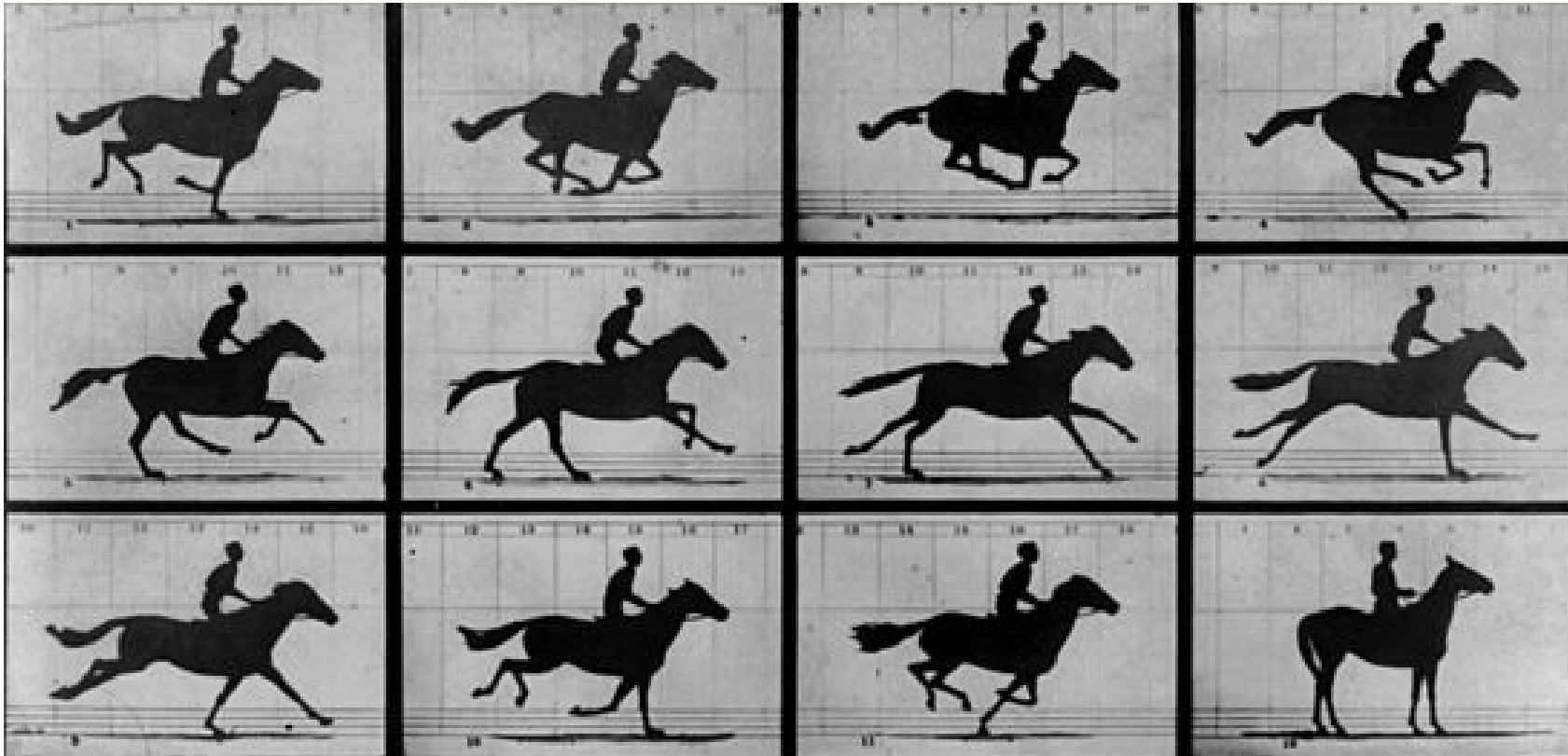# Making and working with Excel chart templates

- **Make your chart**
  - Include as many data series as you think you will ever need
  - In the Chart Tools tab, Design sub-tab, "Save As Template"

- **Note that "Change Chart Type" is also an option**
- **Excel can't read your mind**
  - "Save As Template" can't handle super complex charts

Questions & comments

Horse In Motion, Muybridge (1886)

Contact Information

Garrett Bradford
Cartographer / GIS Analyst

Milliman, Inc.
650 California Street 17th Floor
San Francisco, CA 94108

garrett.bradford@milliman.com
(415) 394-3792

Alice Underwood
Executive Vice President

Willis Re Inc.
200 Liberty Street

New York, NY 10281-1003

alice.underwood@willistowerswatson.com
(212) 915-8439