



Data Visualization in Business Communication

Theory, Methods, and Tools

Gamma Iota Sigma
2023 Regional Conference
Fort Worth, TX

INTRODUCTIONS



Dalesa Bady, ACAS, MAAA
Actuary, GuideOne Insurance

- + Based in Dallas, Texas with over 12 years of experience in the property/casualty insurance industry
- + Areas of work include personal/commercial lines ratemaking, product development, and predictive analytics
- + Passionate about coaching/mentorship
- + Casualty Actuarial Society (CAS) University Liaison for UT Austin/UT Dallas and Chair of International Association of Black Actuaries (IABA) Scholarship Committee



Bryce Chamberlain, ASA, MScA
Principal, Oliver Wyman Actuarial Consulting

- + Associate of the Society of Actuaries
- + Master of Science in Analytics, University of Chicago
- + Leads a team at building business intelligence apps for the web using R Shiny.
- + Passionate about data visualization, user-friendly design, efficiency and flexibility
- + Lead developer for R packages: easyr, hcslim.

AGENDA



1

Why Visualize?

2

How We Think Visually

3

Things To Avoid

4

Problems & Tools to Solve Them

The last slide will show a link to this deck + all the resources mentioned.

Why Visualize?

“

How many of you pay attention to road signs
while driving?

How much harder would it be if road signs were a full body of text?

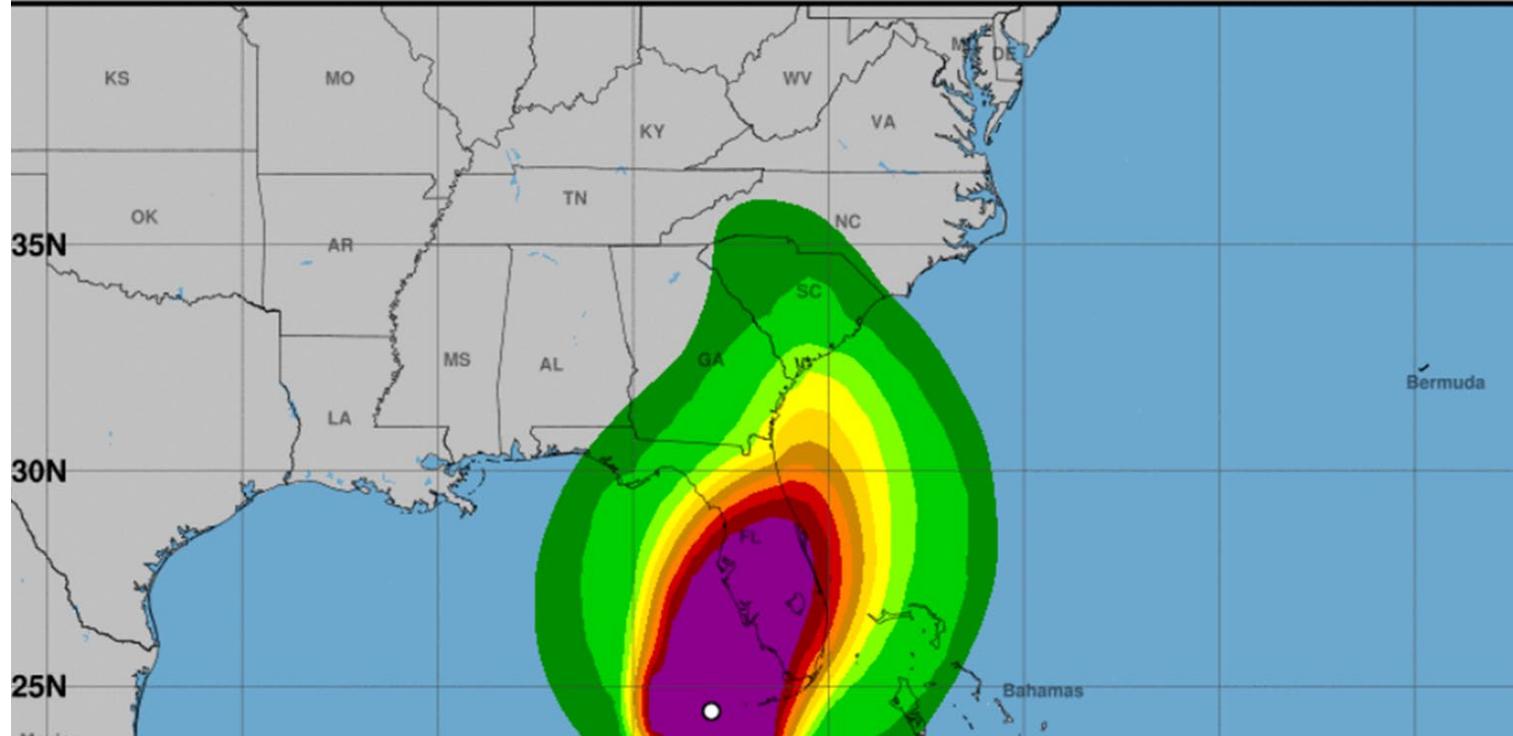


Beware of the road ahead. It might have rained or snowed and it could be slippery, which might cause an incident. Be careful.



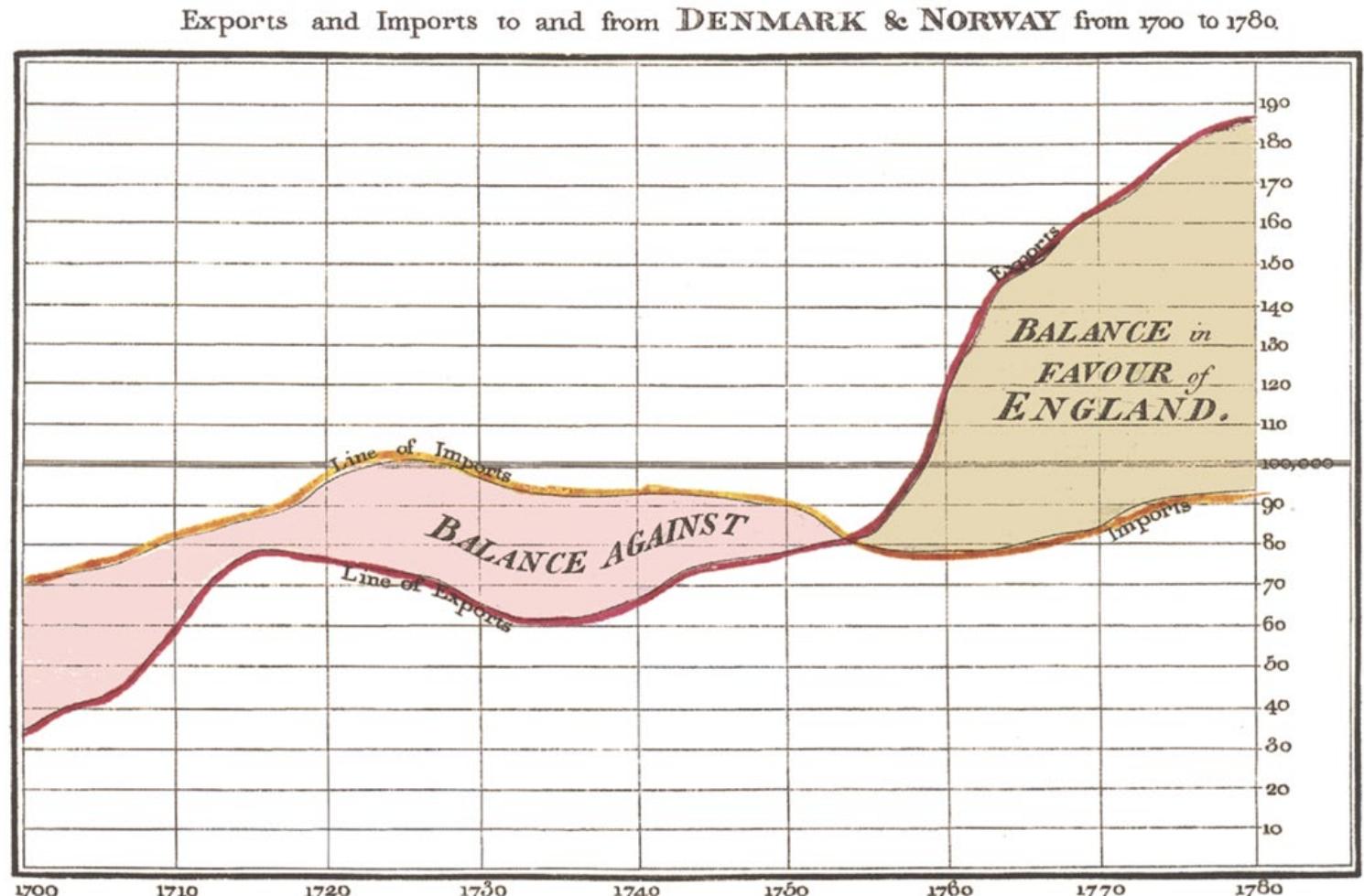
Tropical-Storm-Force Wind Speed Probabilities (Preliminary)

For the 120 hours (5.0 days) from 8 AM EDT WED SEP 28 to 8 AM EDT MON OCT 03



Would this message be as powerful with a table of numbers?

“The Commercial and
Political Atlas”
- William Playfair (1786)



The Bottom line is divided into Years, the Right hand line into £10,000 each.
Published as the Act directs, 1st May 1786. by W^m Playfair
Neale sculpt 352, Strand, London.

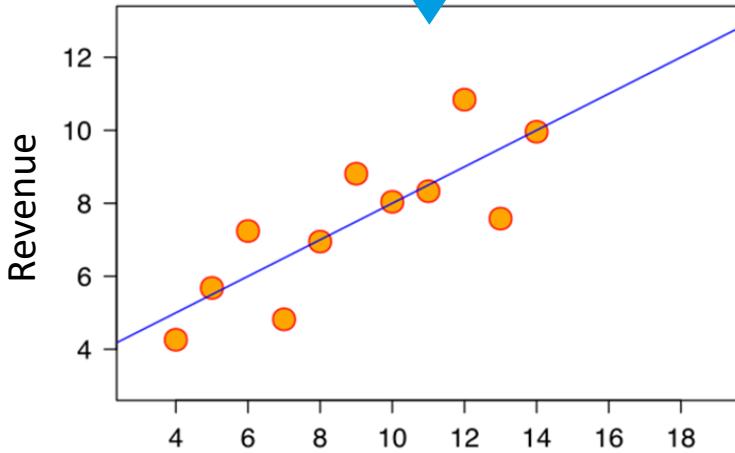
“

Well, just show me the numbers.

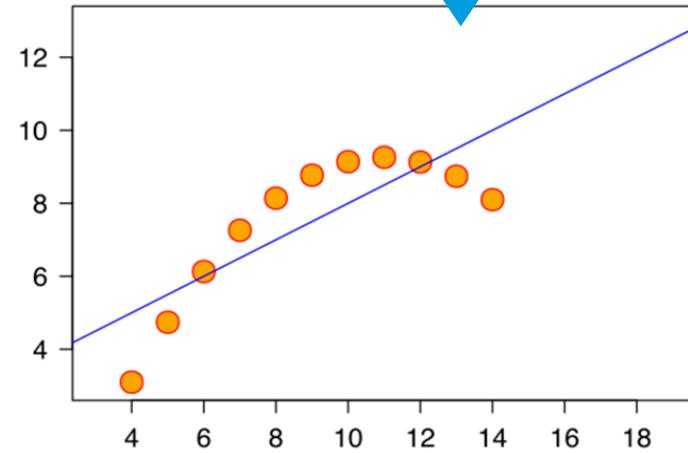
Let's take a look at an example.

Agent	Mean Claims (\$millions)	Mean Revenue (\$millions)	Correlation
A	9	7.5	0.816
B	9	7.5	0.816
C	9	7.5	0.816
D	9	7.5	0.816

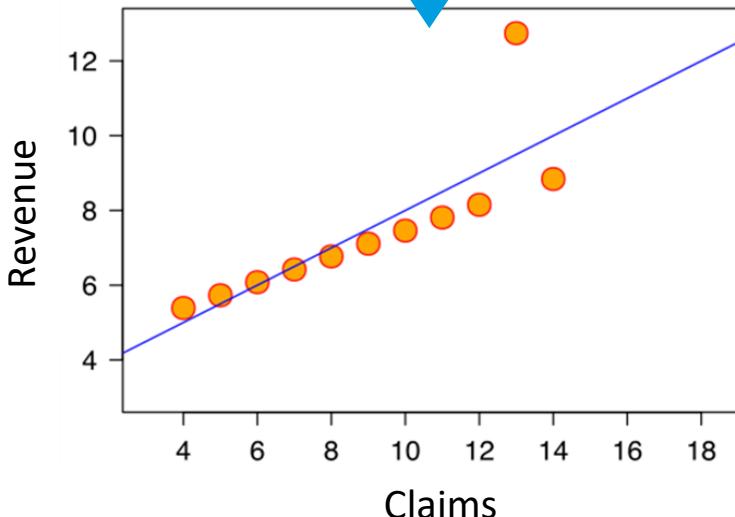
Revenue increases with claims.



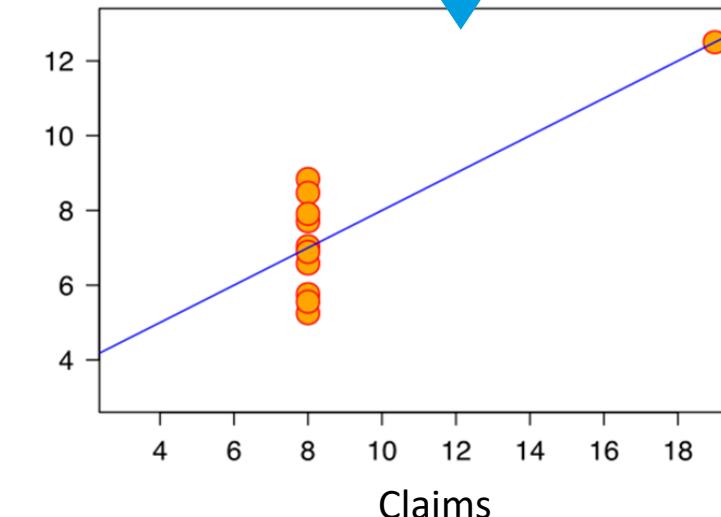
Revenue capped.



Revenue outlier.



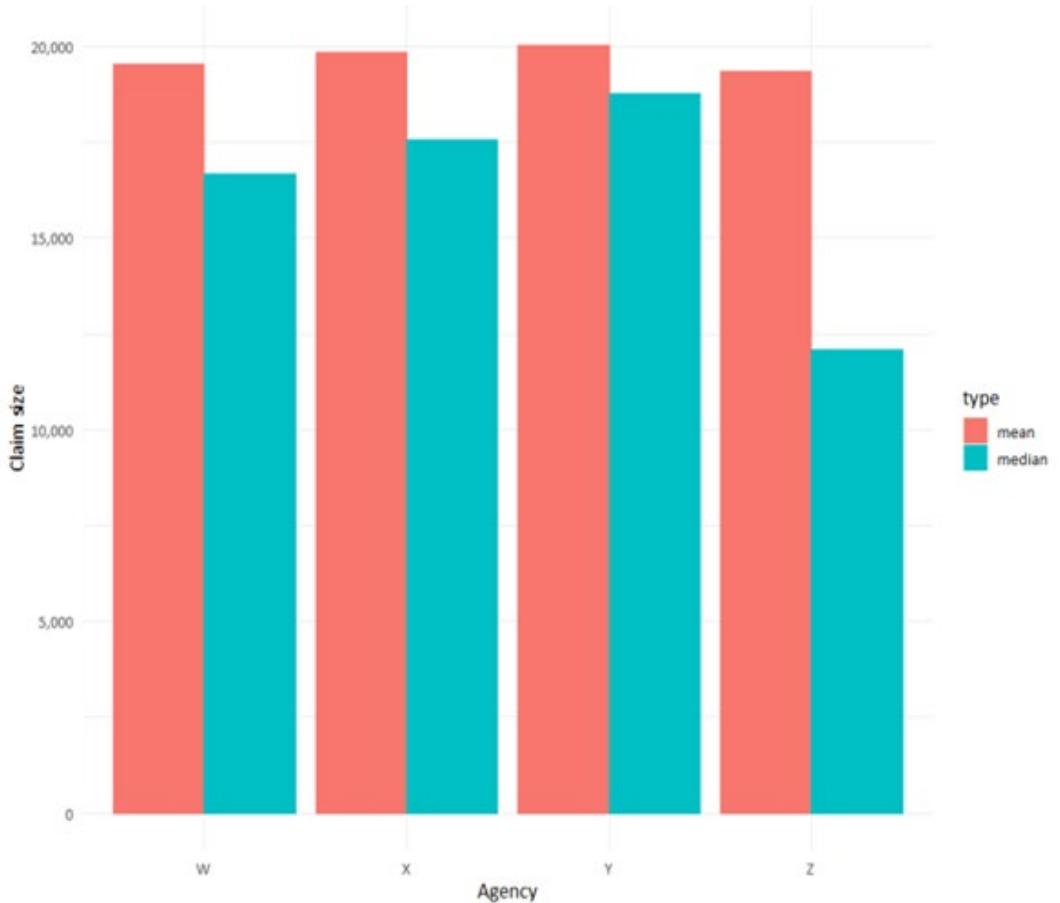
Revenue and amount outlier.



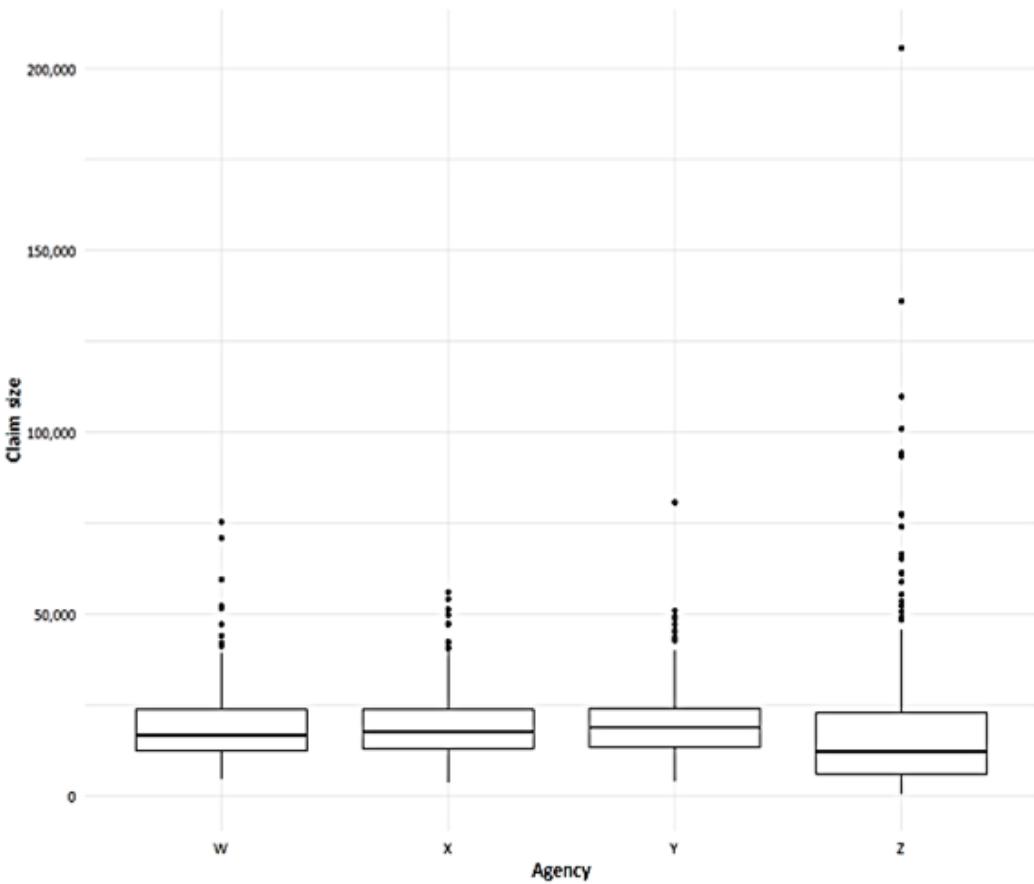
“

Visualization is useful when communicating
to a non-technical person or audience.

Agency	Mean Claim Size	Median Claim Size
W	19,533	16,675
X	19,829	17,577
Y	20,039	18,758
Z	19,363	12,097



Agency	Mean Claim Size	Median Claim Size
W	19,533	16,675
X	19,829	17,577
Y	20,039	18,758
Z	19,363	12,097



How Do We Think Visually?

99999999

9999999999

999,999,999

99,999,999,999

999,999,999

99,999,999,999

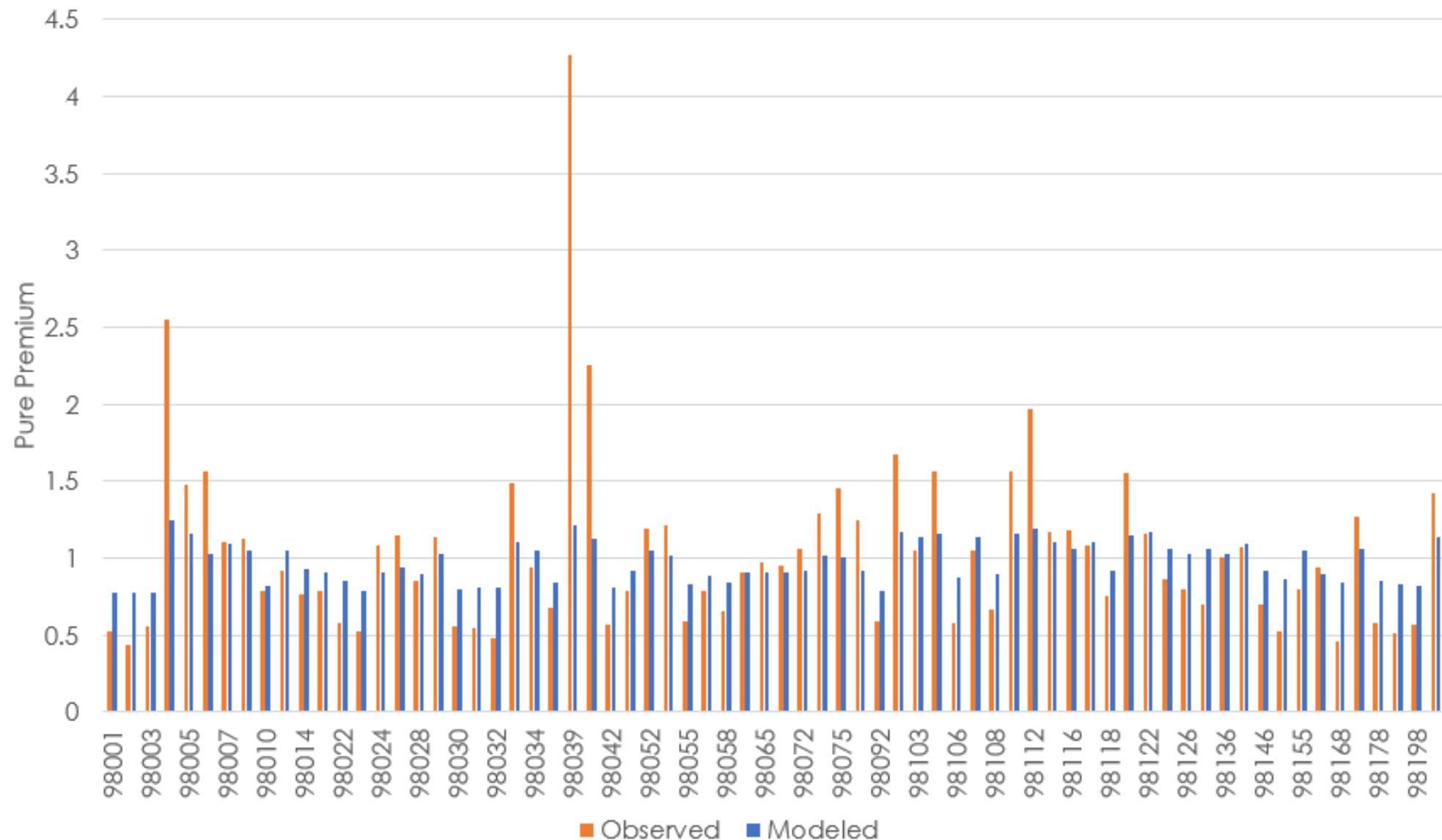


Mapping the values in a visual space helps us understand the enormity of the difference!

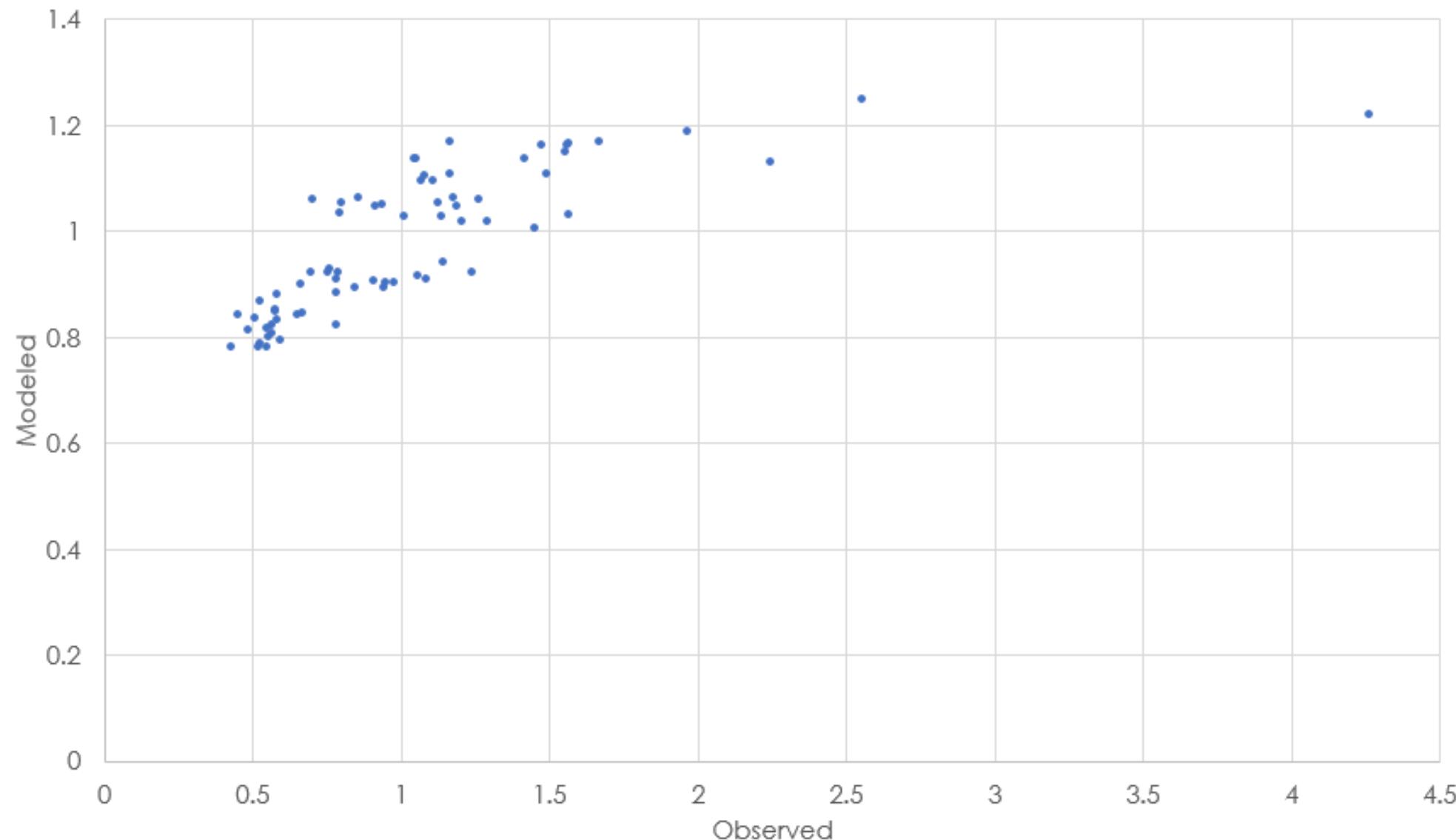
“

Let's take a look at an example of an insurance dataset visualized in two ways.

Observed vs. Modeled Pure Premium by ZIP Code

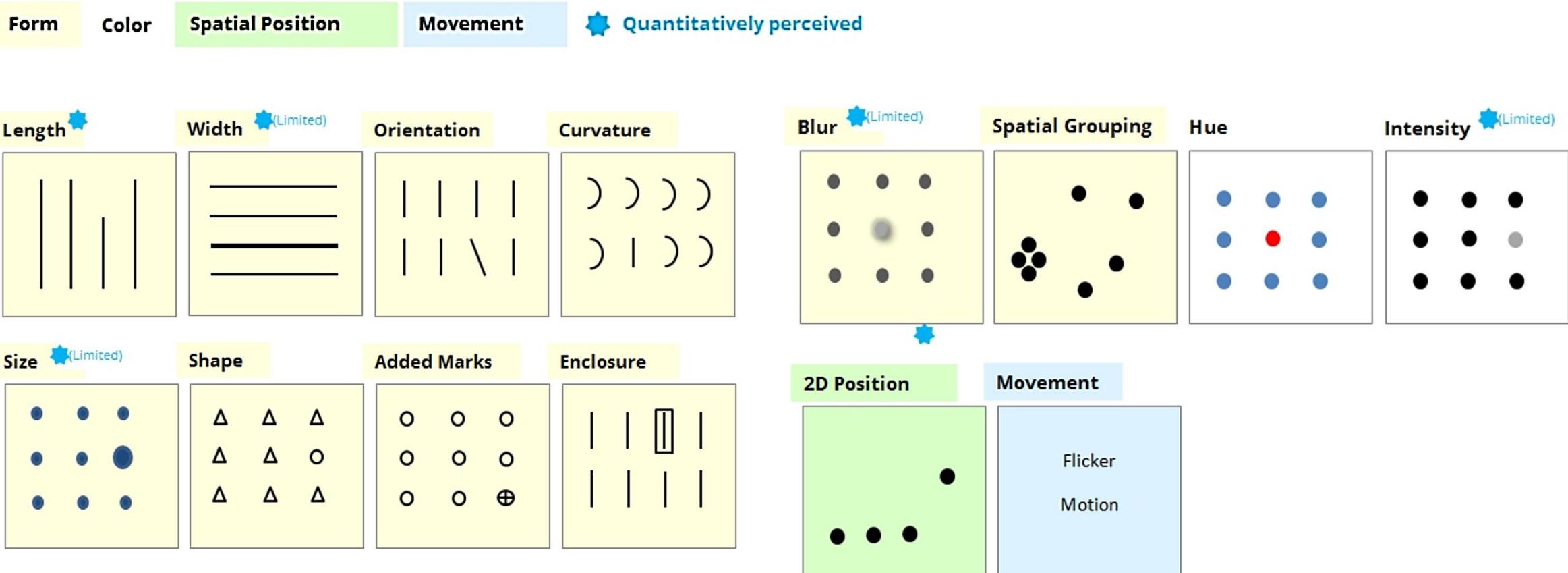


Observed vs. Modeled Pure Premium by ZIP Code

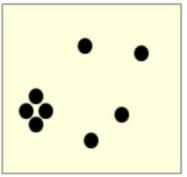


PREATTENTIVE ATTRIBUTES

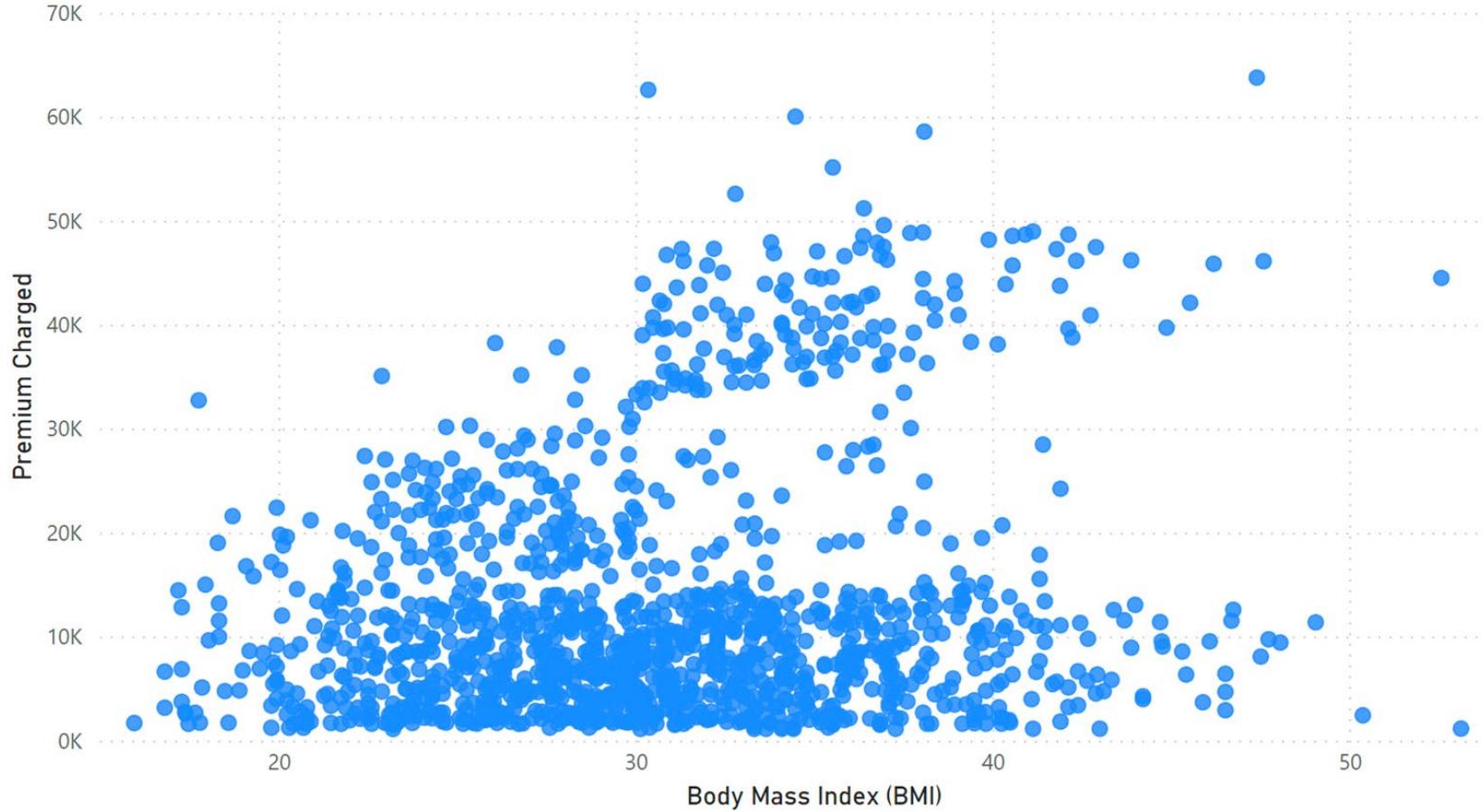
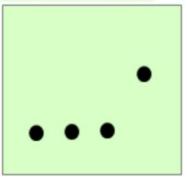
TOOLS FOR COMMUNICATING VISUALLY



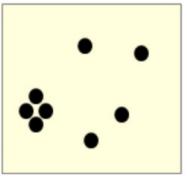
Spatial Grouping



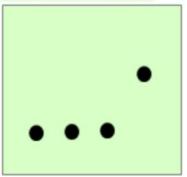
2D Position



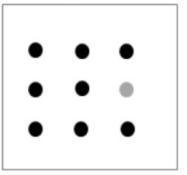
Spatial Grouping



2D Position



Intensity (Limited)



smoker ● no ● yes

70K

60K

50K

40K

30K

20K

10K

0K

Premium Charged

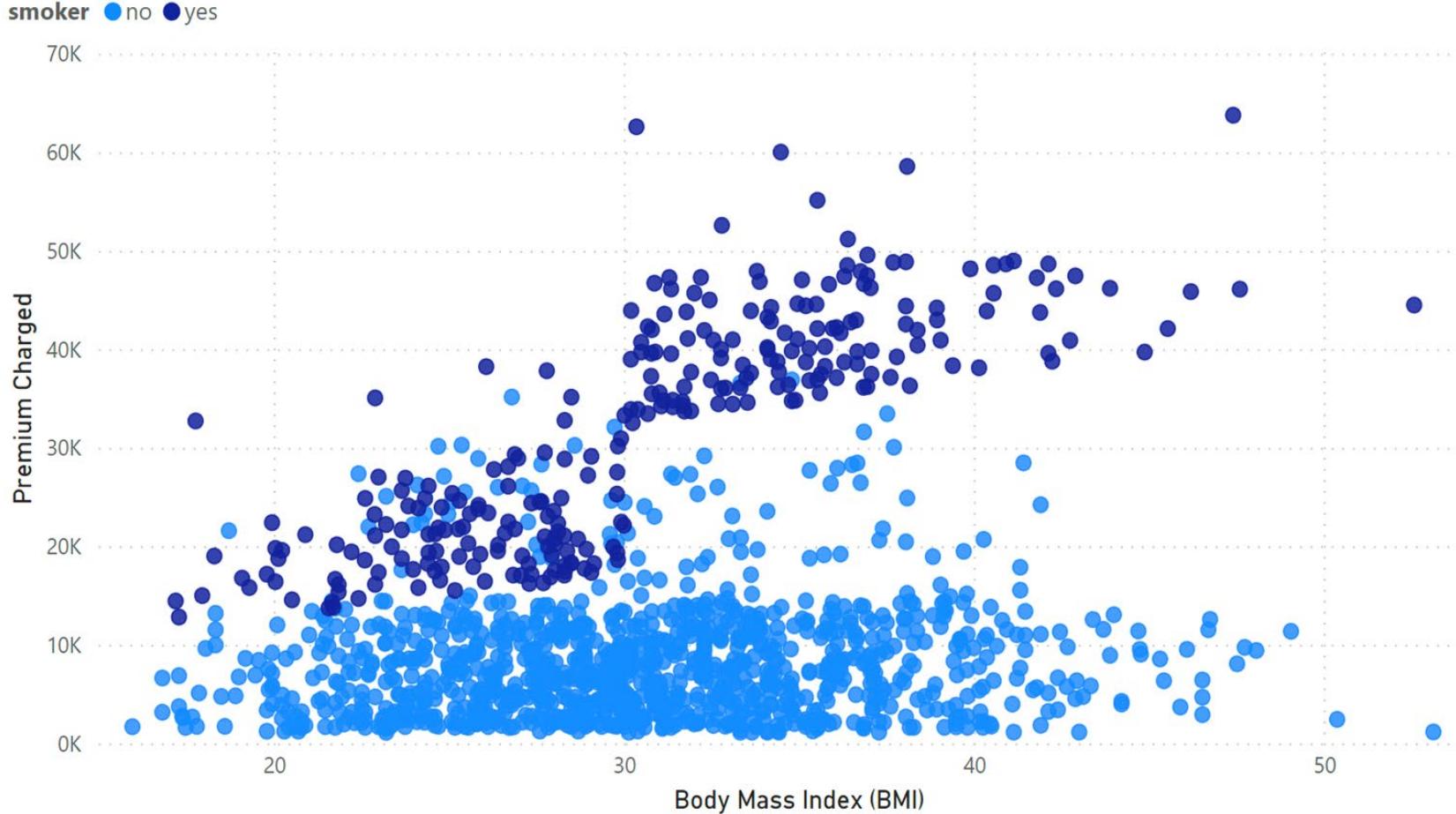
20

30

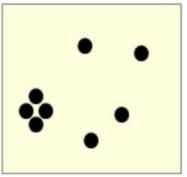
40

50

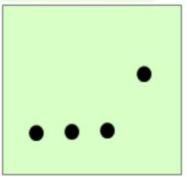
Body Mass Index (BMI)



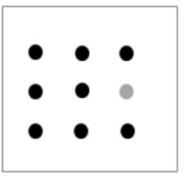
Spatial Grouping



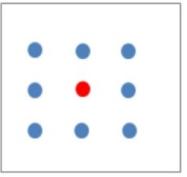
2D Position



Intensity (Limited)



Hue



Sex & Smoker

70K

60K

50K

40K

30K

20K

10K

0K

Premium Charged

70K

60K

50K

40K

30K

20K

10K

0K

Body Mass Index (BMI)

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

30

40

50

20

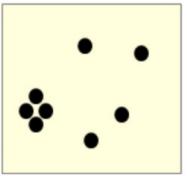
30

40

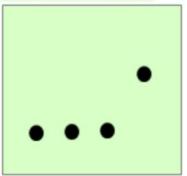
50

20

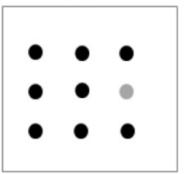
Spatial Grouping



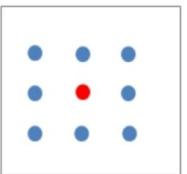
2D Position



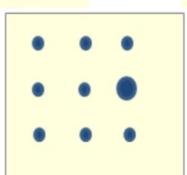
Intensity (Limited)



Hue



Size (Limited)



Sex & Smoker

● female-no ● female-yes ● male-no ● male-yes

60K

50K

40K

30K

20K

10K

0K

Premium Charged

Body Mass Index (BMI)

15

20

25

30

35

40

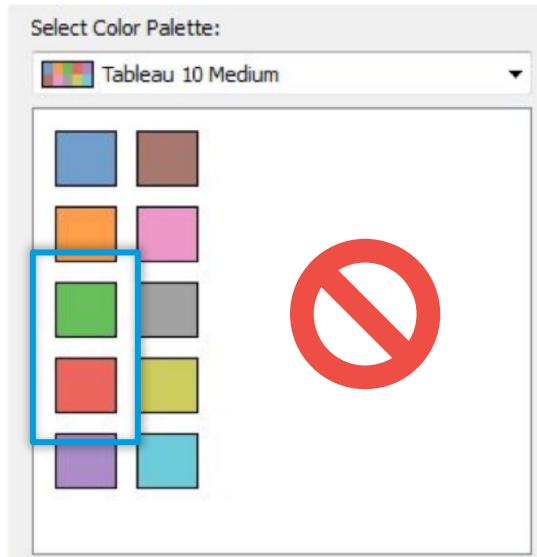
45

50

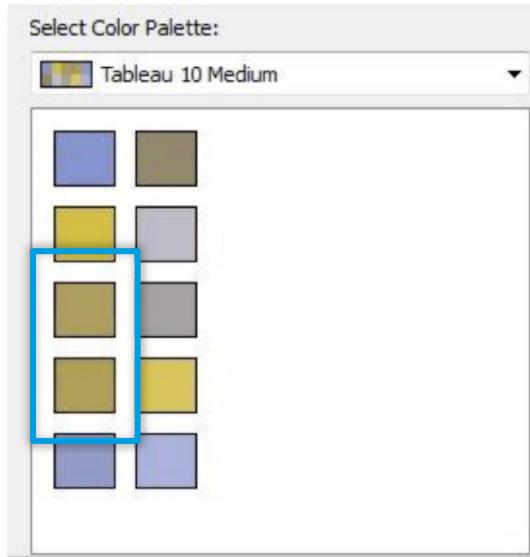
Things to avoid

Red/Green Deutanopia (6% of males)

Original Image

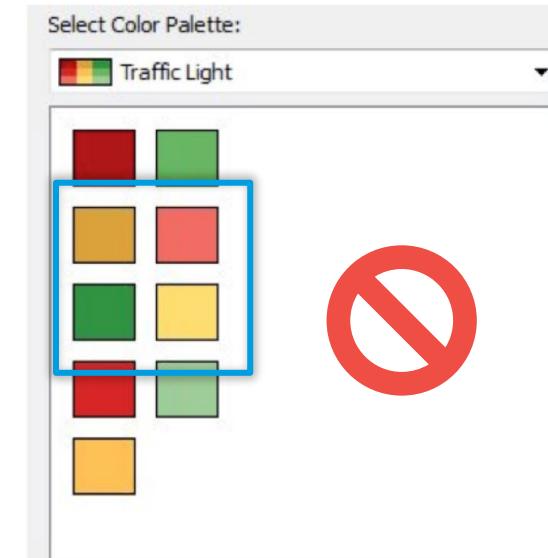


Deutanope Simulation

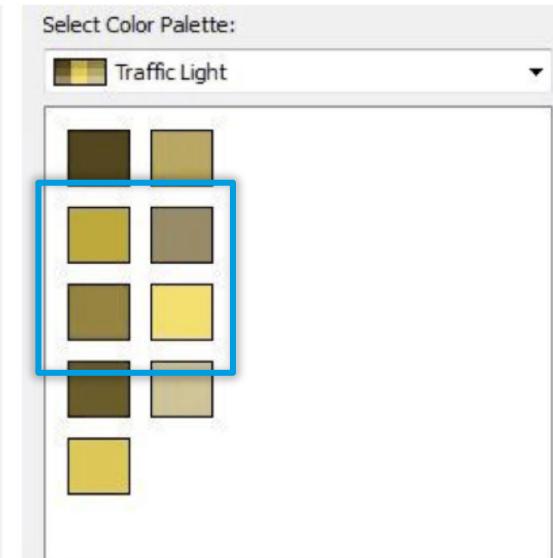


Red/Green Protanopia (2% of males)

Original Image

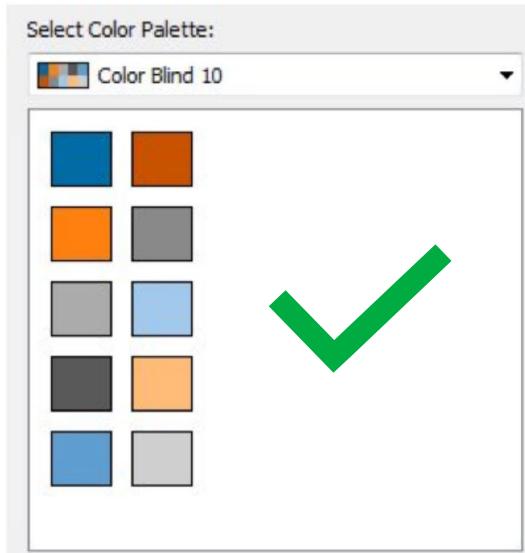


Protanope Simulation

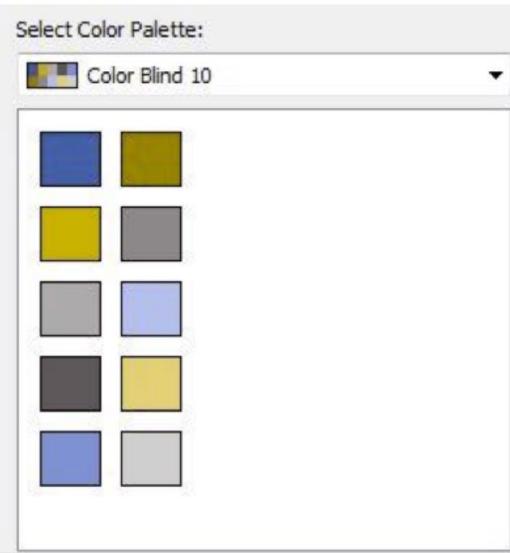


Red/Green Deutanopia (6% of males)

Original Image

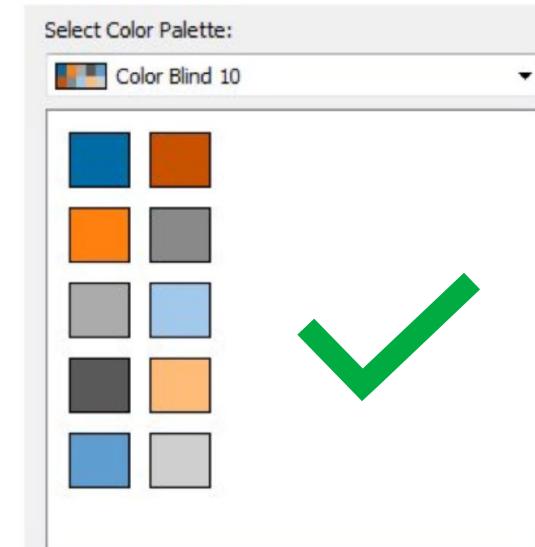


Deutanope Simulation

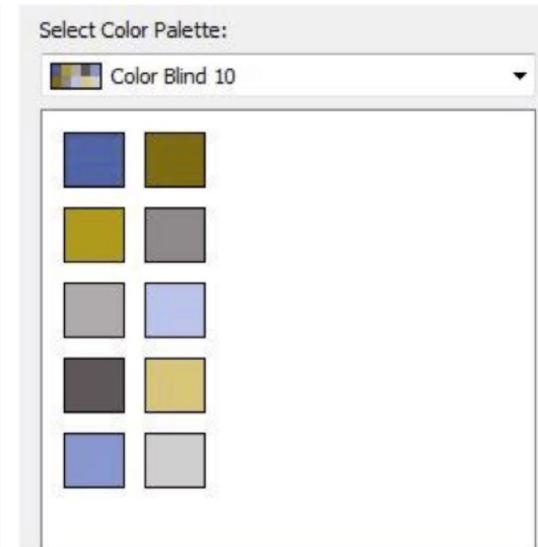


Red/Green Protanopia (2% of males)

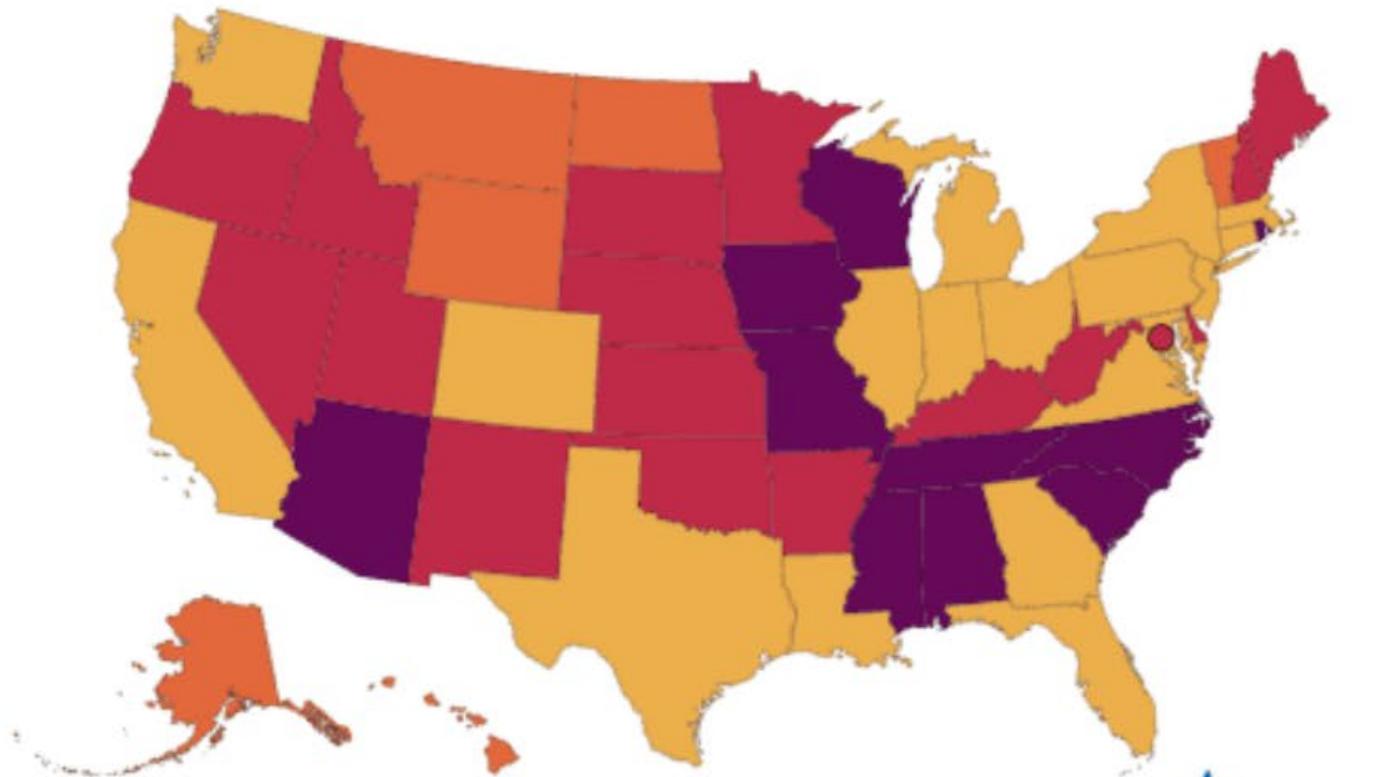
Original Image



Protanope Simulation



18 states report more than 10,000 cases of COVID-19.



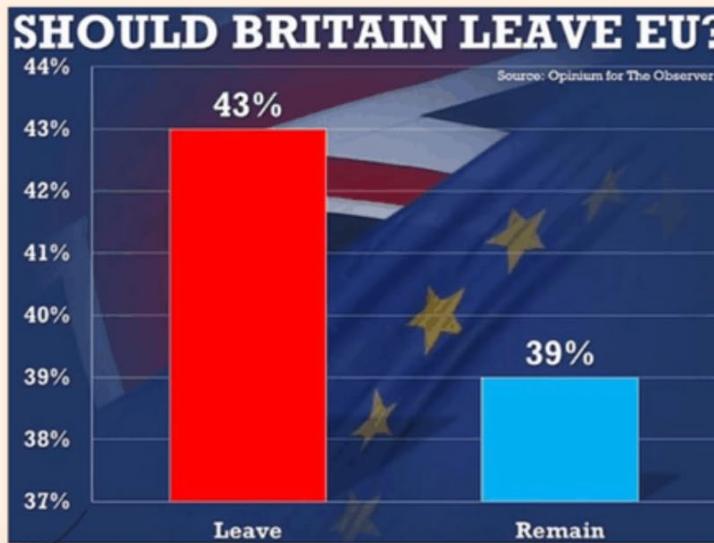
Reported Cases

- | | |
|------------------|-------------------|
| ● 1 to 100 | ● 10,001 or more |
| ● None | ● 101 to 1,000 |
| ● 1,001 to 5,000 | ● 5,001 to 10,000 |



Graphics that
are accurate
but misleading

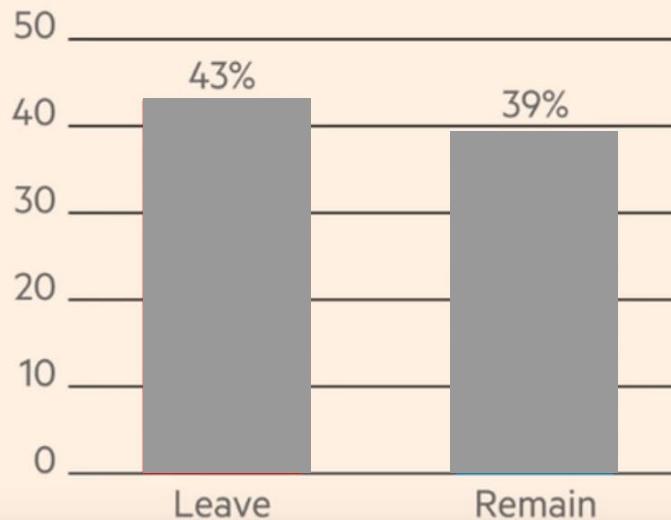
Baseline should
start at zero,
not 37



Graphics that
are accurate
but misleading

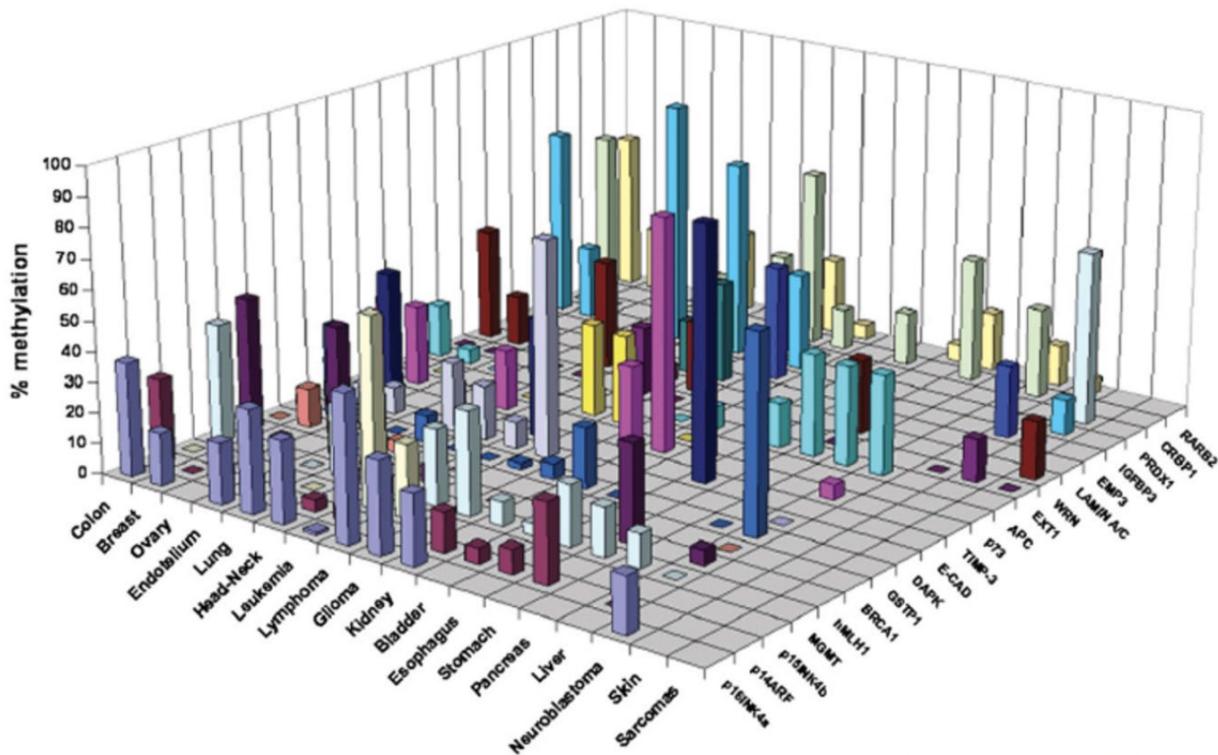
A better chart of
the same data

Should Britain leave the EU?

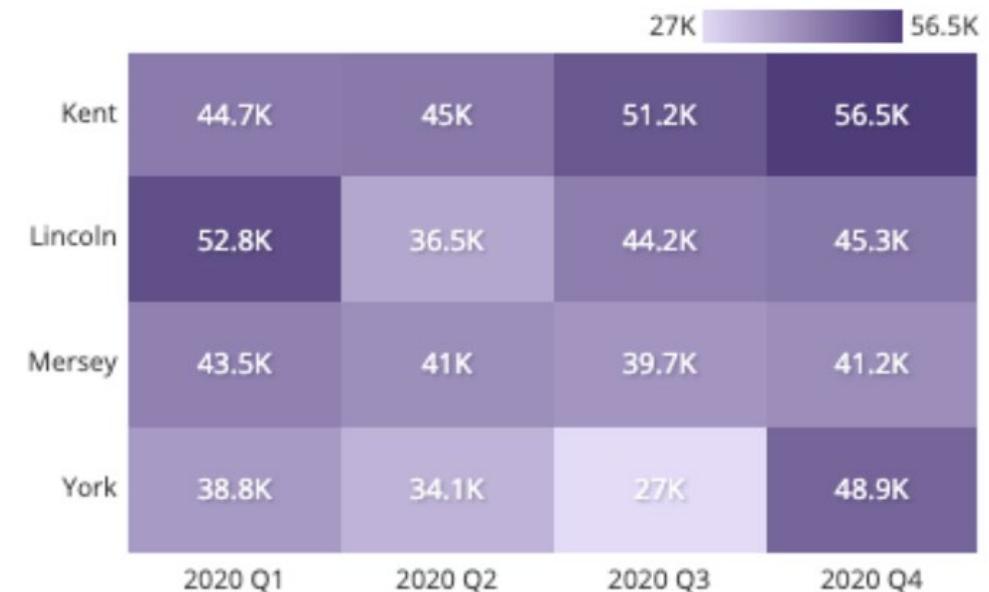




A CpG Island Hypermethylation Profile of Human Cancer



New Revenue



Problems & Tools to fix them

I HAVE DATA, BUT I DON'T KNOW WHAT STORY TO TELL.

- Build a predictive model and look at important features.
- Do this quickly with an automated machine learning tool: RapidMiner (point/click), storyteller (code/R), etc.

```
# run model a model to find a story about charges (premiums)
dt %>%
  correlatedfeatures_address(
    target = 'charges'
  ) %>%
  fitmodel() %>%
  summary()

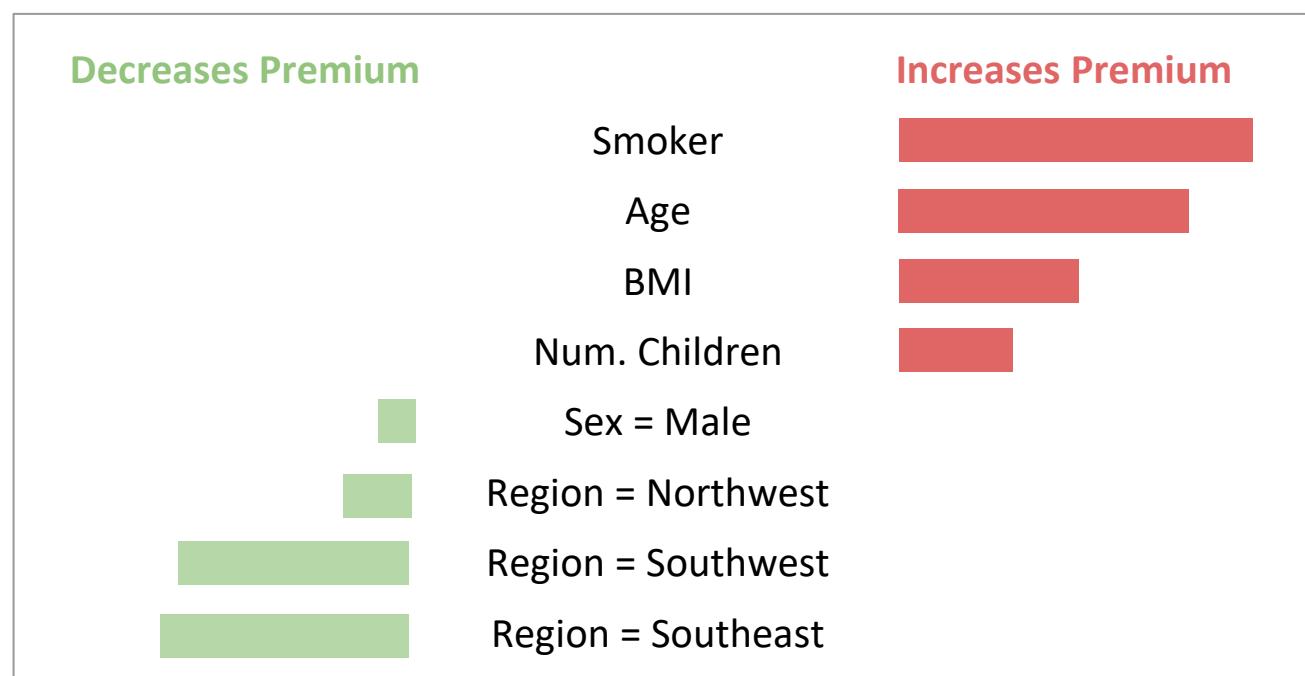
[1] "charges"

Call:
lm(formula = y ~ ., data = yX)

Residuals:
    Min      1Q  Median      3Q     Max 
-10584 -2748 -1068   1092  24373 

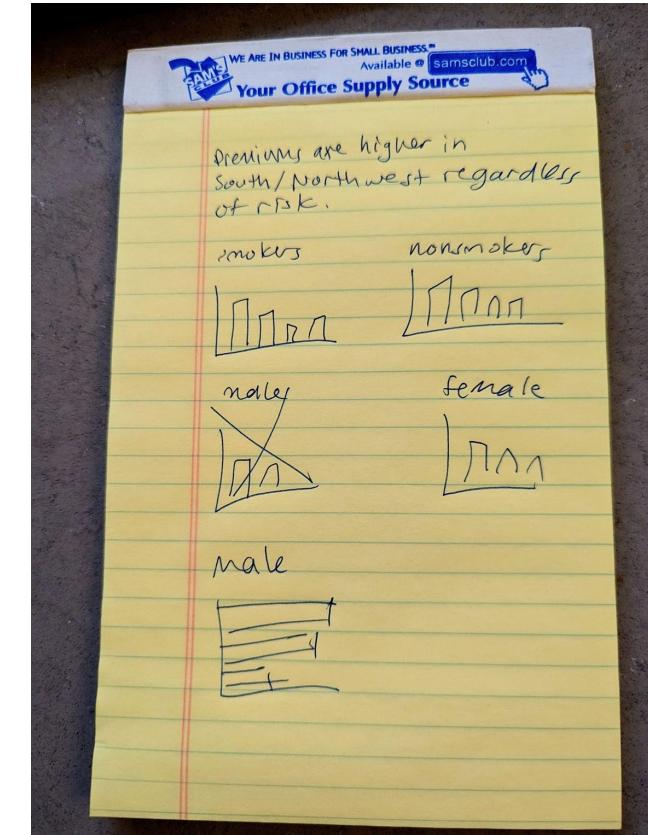
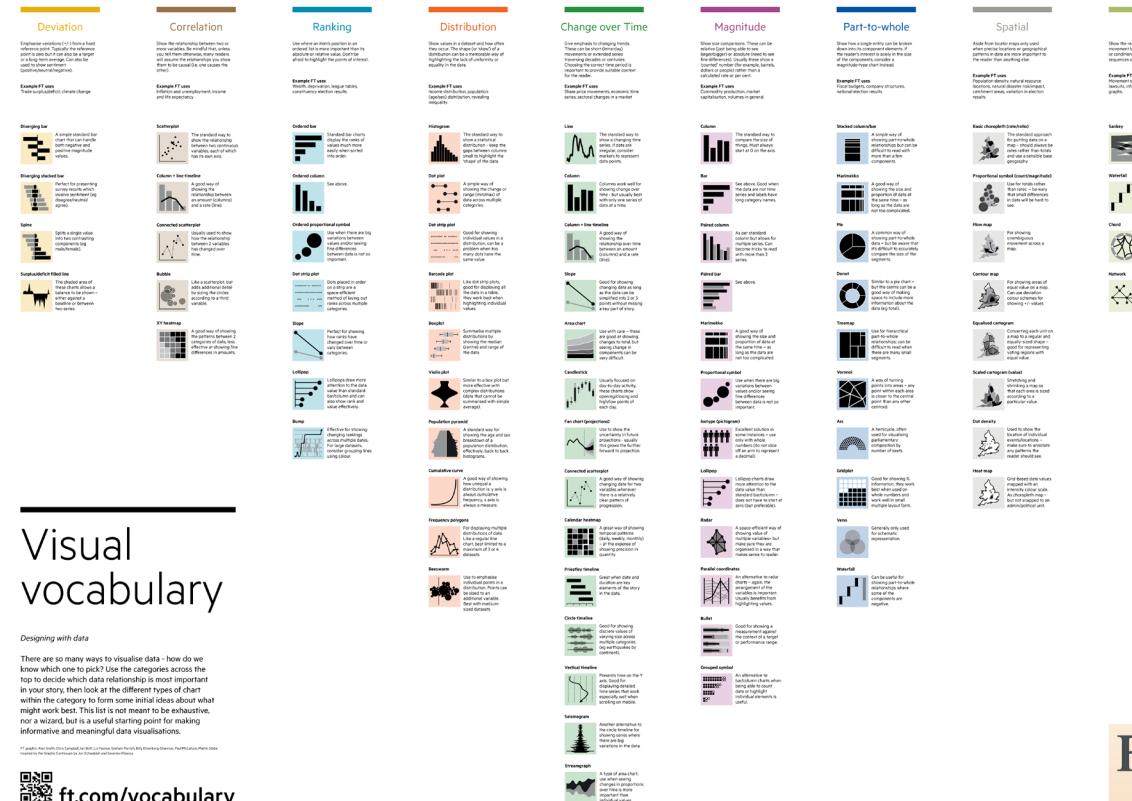
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -11003.70    965.52 -11.397 < 2e-16 ***
age          250.67     11.48  21.841 < 2e-16 ***
bmi          317.37     28.15  11.273 < 2e-16 ***
children     519.40    132.24   3.928 9.03e-05 ***
smokerTRUE   22885.98   403.69   56.692 < 2e-16 ***
sex.male     -106.99    320.02  -0.334  0.7382  
region.southeast -1072.18  460.65  -2.328  0.0201 *  
region.northwest -444.34  456.46  -0.973  0.3305  
region.southwest -1021.20  458.71  -2.226  0.0262 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5787 on 1309 degrees of freedom
Multiple R-squared:  0.7448,    Adjusted R-squared:  0.7432 
F-statistic: 477.4 on 8 and 1309 DF,  p-value: < 2.2e-16
```



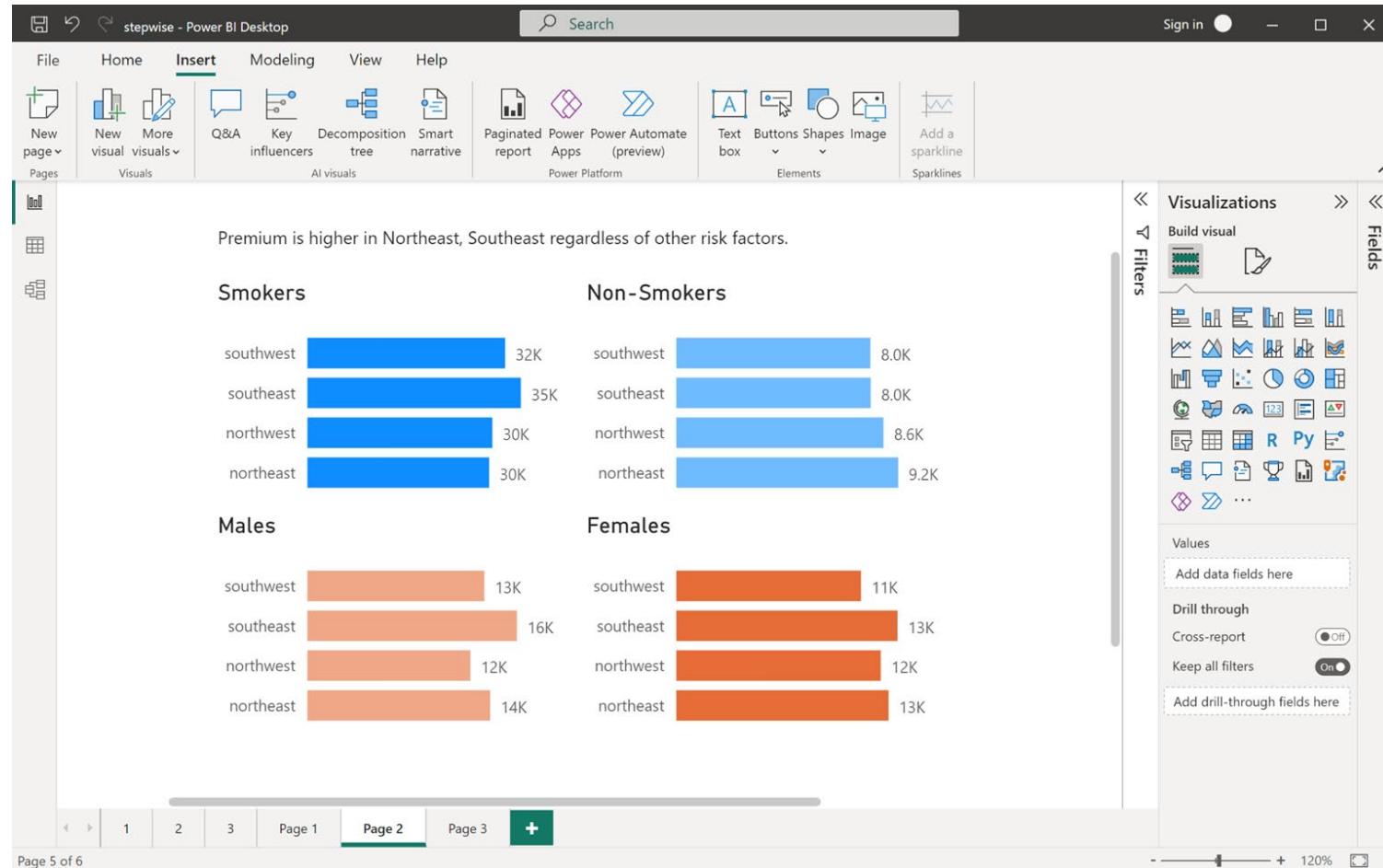
I KNOW THE STORY I WANT TO TELL, BUT WHAT VIZ SHOULD I USE?

- Get inspiration from the Financial Times Visual Vocabulary
- Use hand-drawing to quickly prototype an idea.



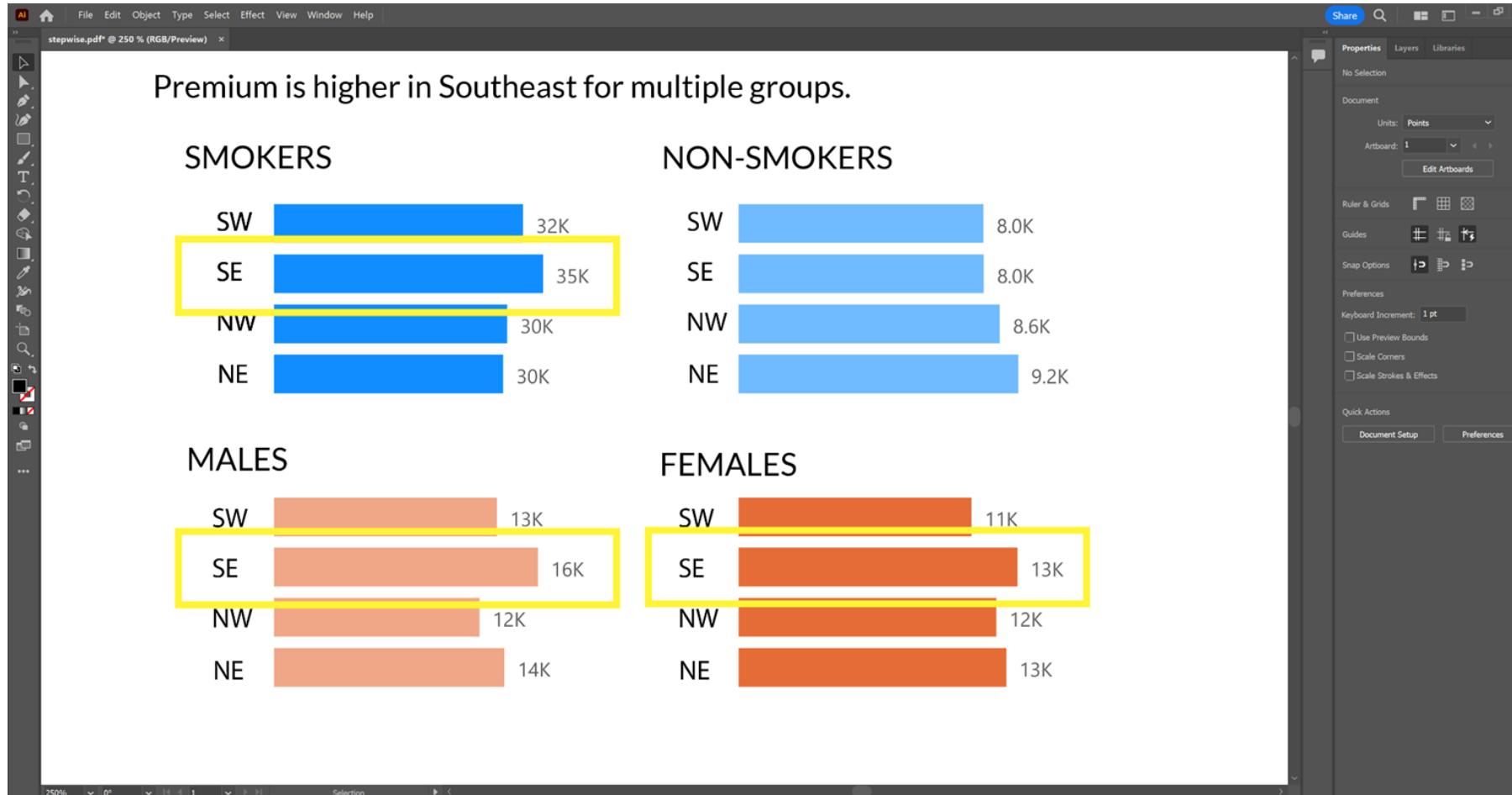
I AM READY TO MAKE A VISUALIZATION, BUT I AM SHORT ON TIME.

- Use a click/drag tool: PowerBI Free Desktop, Excel PivotChart, etc.
- Avoid code-based solutions that can take a long time.



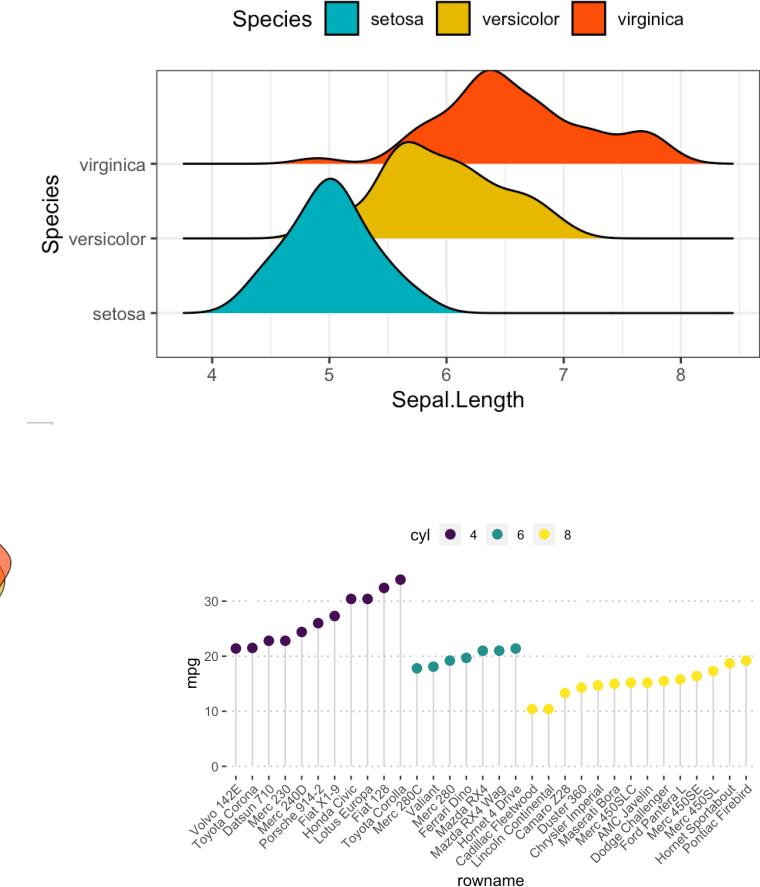
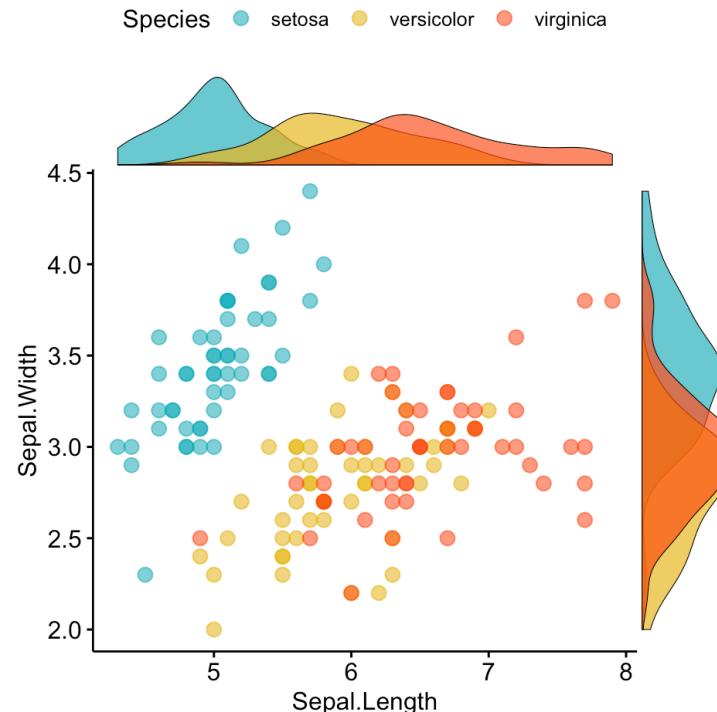
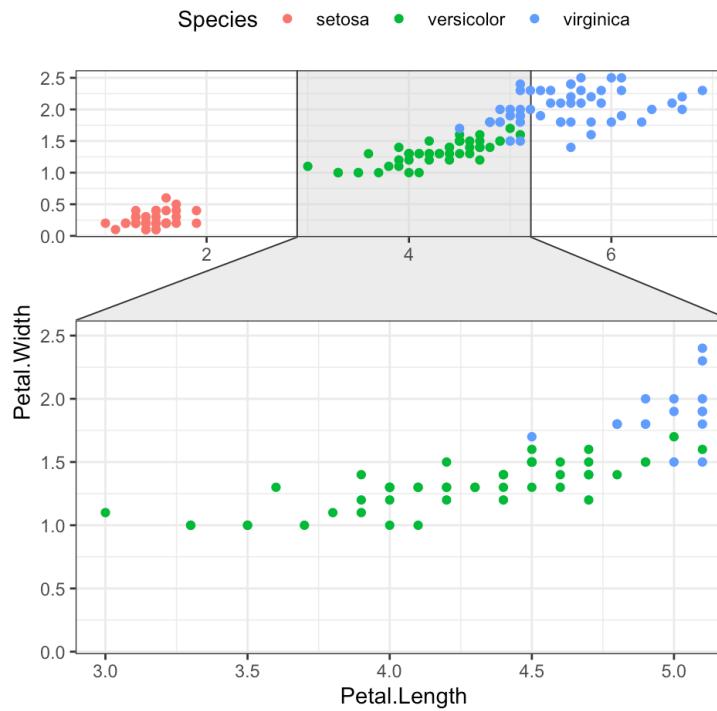
I'M NOT HAPPY WITH THE DESIGN OPTIONS IN POWER BI

- Export from GUI to PDF and edit with a design tool like Adobe Illustrator, Inkscape, or Fiji.
- This is not intuitive, so watch a video to see how (link on last slide).



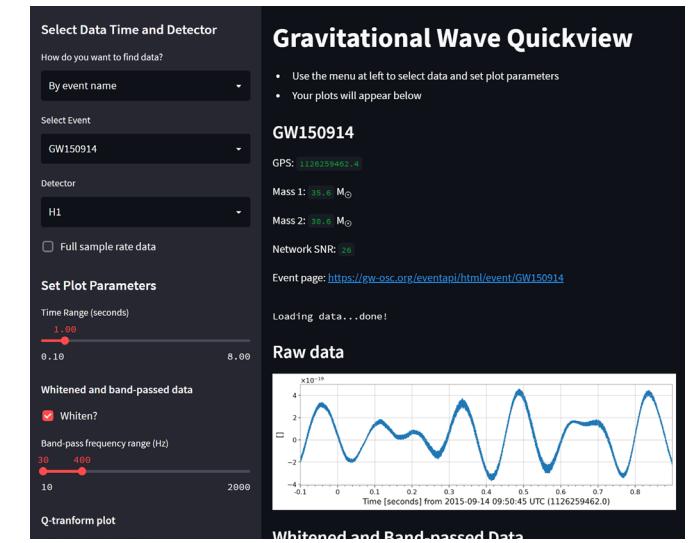
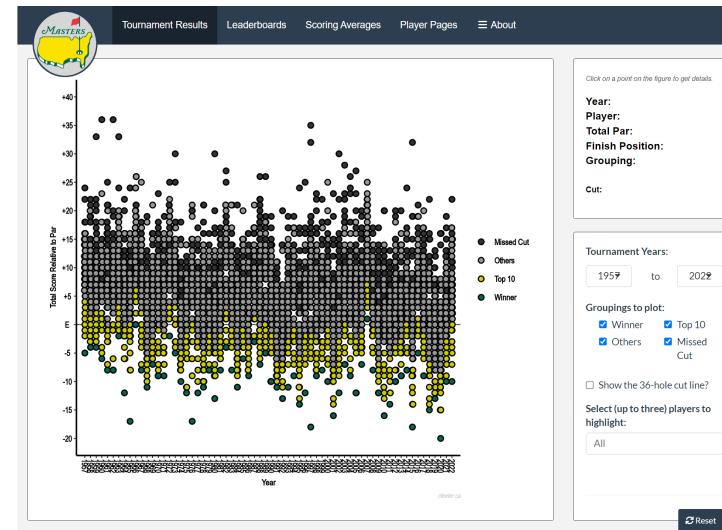
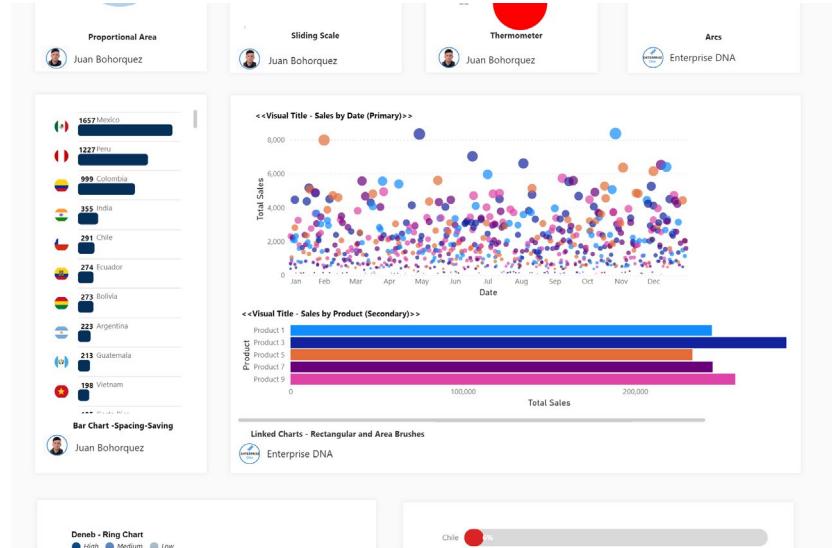
I NEED TO MAKE MANY SIMILAR CHARTS / NOT HAPPY WITH CHART OPT

- Use a code-based solution so you can automate the process or use new chart types.
- R: ggplot, RMarkdown; Python: seaborn, Jupyter.



I NEED TO BUILD A DYNAMIC APPLICATION.

- Use a tool you can publish to the web and give users power to filter, etc.
- Power BI, Tableau (click/drag), R Shiny (code, R) or Streamlit (code, Python)



<https://community.powerbi.com/t5/Data-Stories-Gallery/My-own-Gallery/td-p/3054132>

<https://shiny.rstudio.com/gallery/masters.html>

<https://gw-quickview.streamlit.app/>

AutoML

Let the computer do the work for you.
Rapidminer, Storyteller

Found your story?

Free-hand Drawing

Quickly prototype your final product.

Know what you need?

Business Intelligence

Prototype with data.
PowerBI Desktop, Excel PivotChart

Extra mile,
one time?

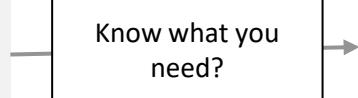
Design

Adobe Illustrator

Need to automate?

Code-based

R, R Shiny, Python



THANK YOU



Dalesa Bady, ACAS, MAAA
Actuary, GuideOne Insurance
linkedin.com/in/dalesa-bady-acas-05822336



Bryce Chamberlain, ASA, MSCA
Principal, Oliver Wyman Actuarial Consulting
linkedin.com/in/brycechamberlain

Get this deck + linked resources at
<https://bit.ly/gis23-dataviz>



Worksheets

Calculator
Excel, Google Sheets

Most people can use it.
Easily manipulate single records.

Easy to make mistakes.
Difficult to automate.
Slow on large data.

Business Intelligence

Get started now
Power BI, Tableau, PivotCharts

Lots of options quickly.
Click & drag

Limited functionality.
Difficult to automate.

AutoML

Search for insights
RapidMiner, Storyteller

Find stories across all data.

Limited visualization.
Results are complex.
Expensive if not open source.

Design

Make it pretty
Adobe Illustrator, Inkscape

Lots of features and options for perfecting the visual.

Very time consuming.
Software is complex, difficult to learn.

Free-hand Drawing

Begin with the end in mind
Pen & Paper, Tablet

Fastest method, no interface to slow you down.

Not generated by data.
Not fit for delivery.

Code-based

Automate repetitive tasks
RMarkdown, R Shiny

Unlimited functionality.
Open source.
Git version control.

Time consuming.
Need to code.