Be a *Visuary*

Data Visualization for Actuaries

2025.02.11

Enbo Jiang, FCAS, CERA

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.

Disclaimers

- The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the official policy or position of the presenter's employer or the Casualty Actuarial Society. Any opinions, findings, conclusions, or recommendations expressed in this presentation are those of the presenter alone and do not necessarily reflect the views of the presenter's employer, the Casualty Actuarial Society, or any other organization or entity.
- This presentation is for informational purposes only and does not constitute professional advice. While every effort has been made to ensure the accuracy and completeness of the information provided, the presenter makes no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability with respect to the content of this presentation. Any reliance you place on such information is therefore strictly at your own risk.
- The presenter shall not be liable for any loss or damage arising from the use of, or reliance on, the information contained in this presentation, including, but not limited to, indirect or consequential loss or damage, or any loss or damage whatsoever arising from loss of data or profits arising out of, or in connection with, the use of this presentation.

Visuary

[noun]

An actuary who is *thoughtful* about communicating the numeric findings of actuarial analyses in visual forms.

Poll: What is your practice area?

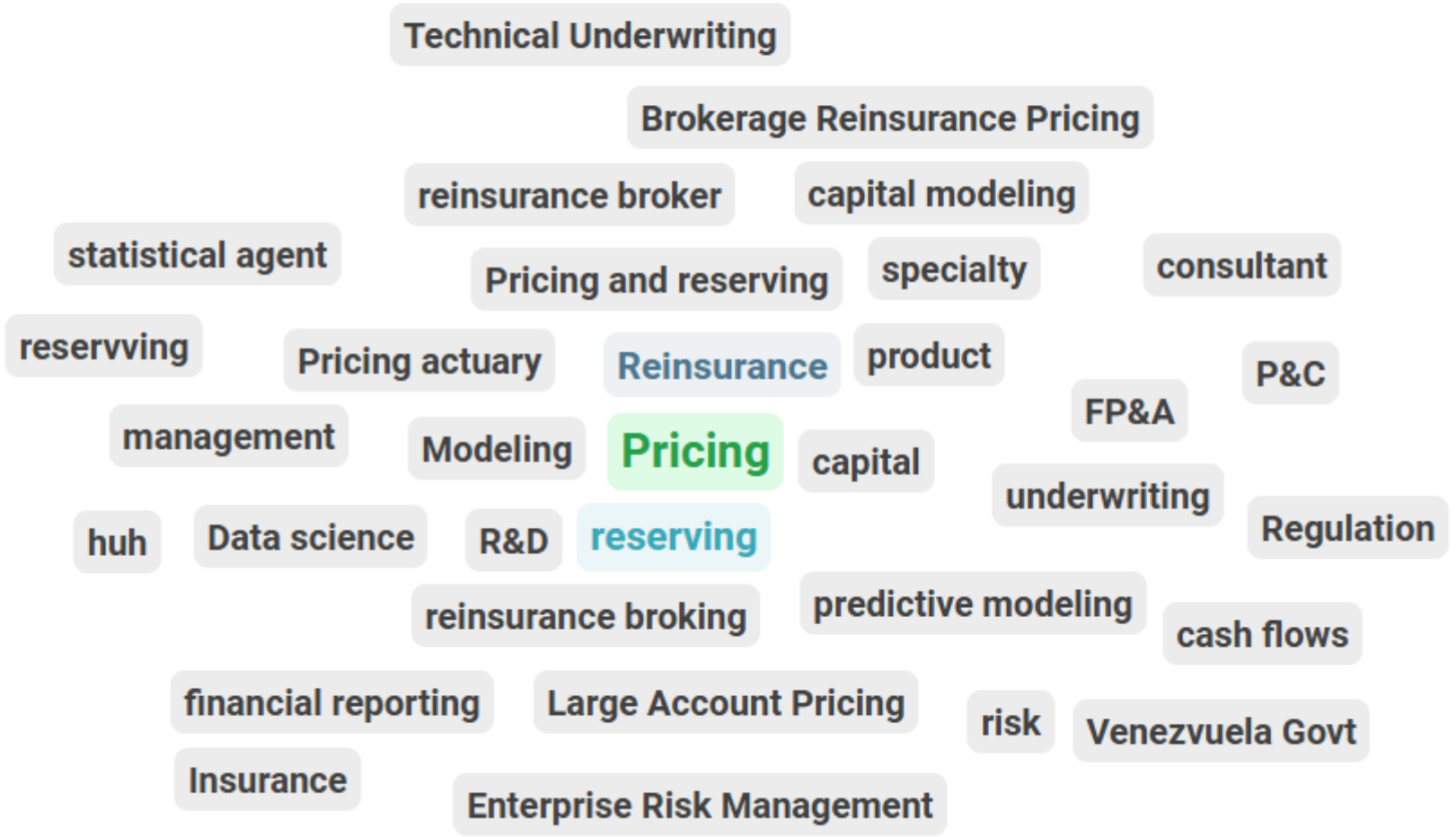
Word cloud
288 votes

Poll settings

What is your practice area? Pricing, reserving, reinsurance, capital, product, underwriting, etc.

Review answers

- Pricing 135
- reserving 50
- Reinsurance 20
- Pricing and reserving 8
- Modeling 8
- Pricing actuary 7
- capital 6
- predictive modeling 4
- product 4

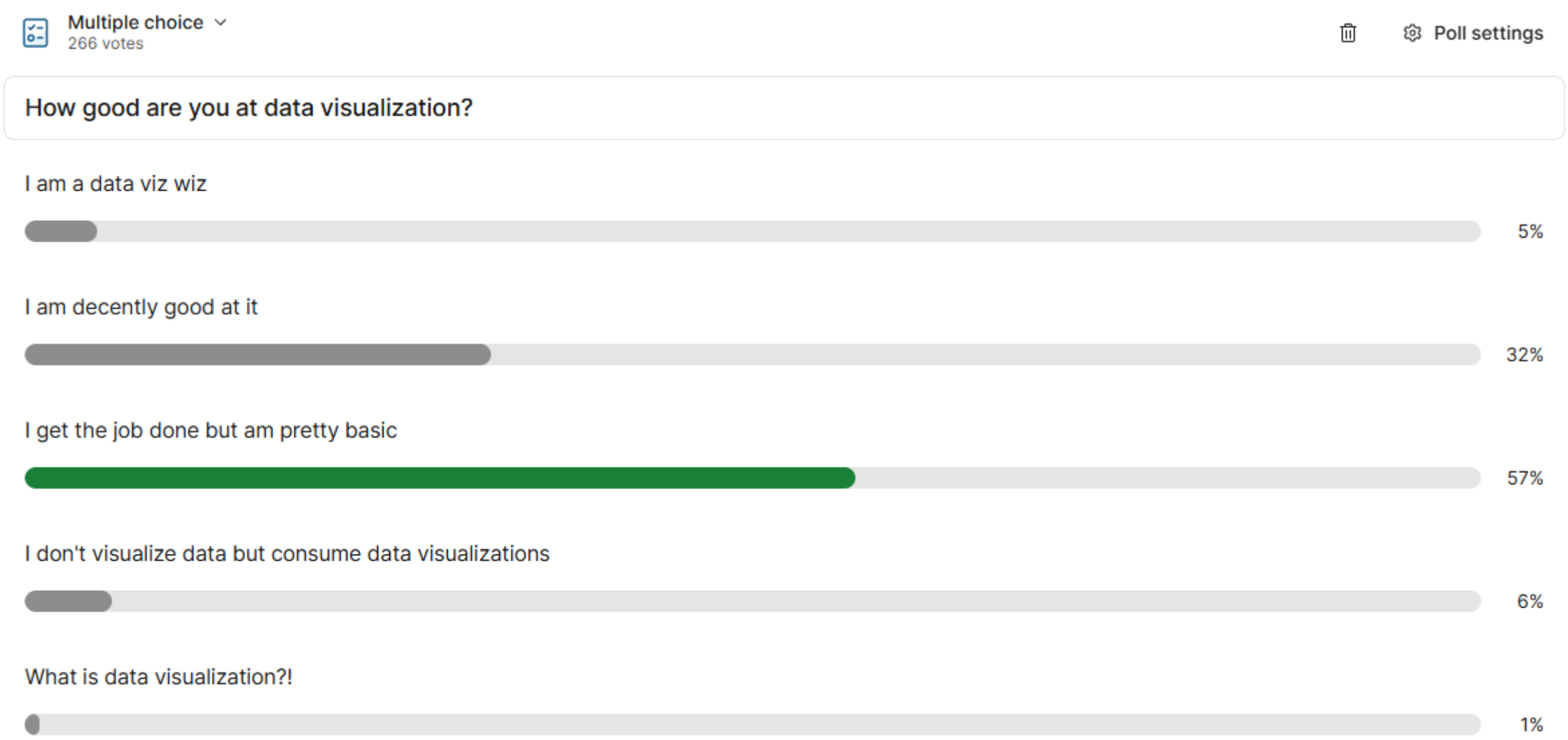


Agenda

- Introduction
- Theories and Principles
- Applications
  - Design Execution
  - Contextual Awareness

Poll: How good are you at data visualization?

A. I am a data viz wiz
B. I am decently good at it
C. I get the job done but am pretty basic
D. I don’t visualize data but consume data visualizations
E. What is data?!


Multiple choice
266 votes
Poll settings
How good are you at data visualization?
I am a data viz wiz
5%
I am decently good at it
32%
I get the job done but am pretty basic
57%
I don't visualize data but consume data visualizations
6%
What is data visualization?!
1%

Poll: What is your go-to data visualization tool?

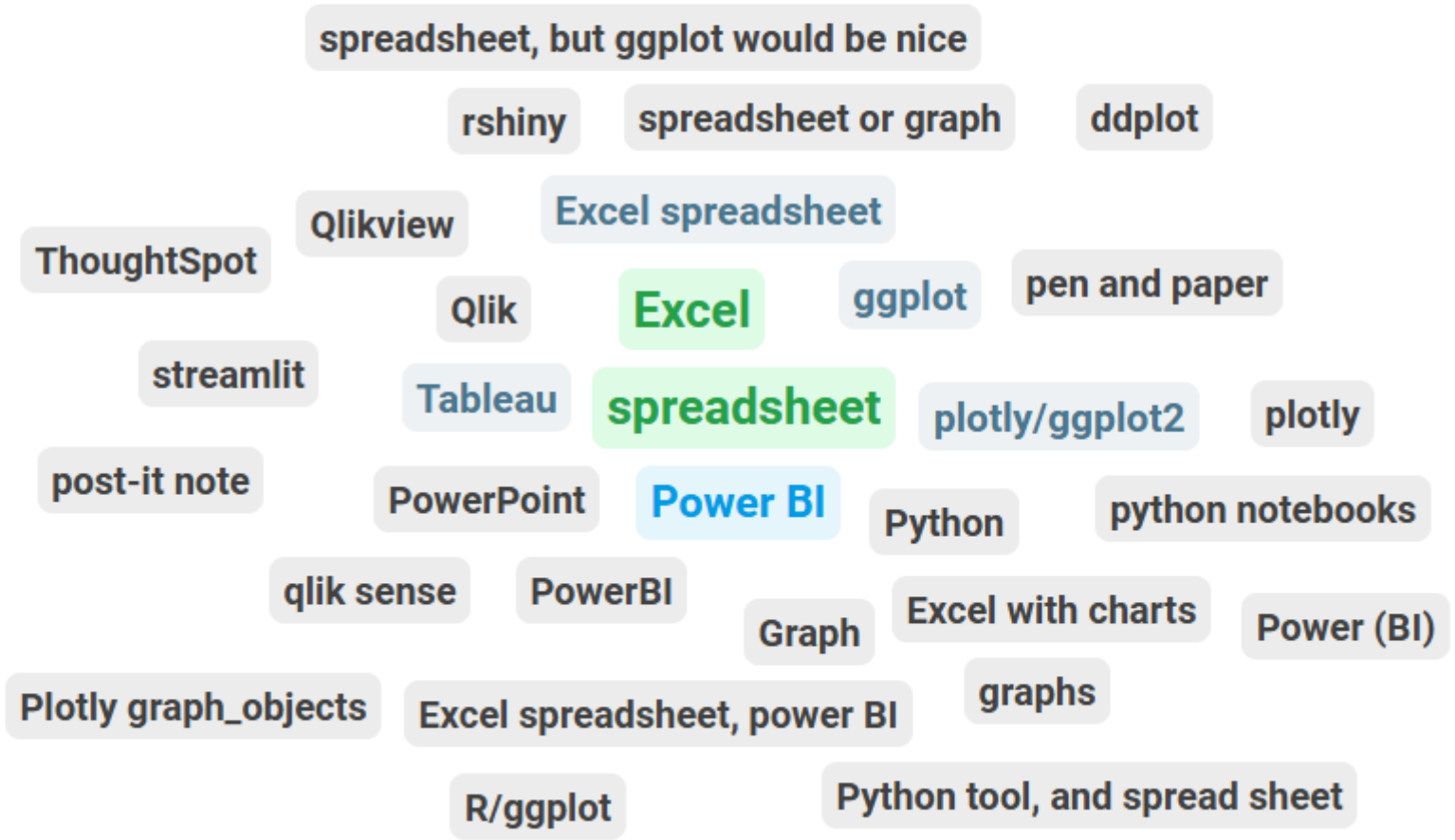
Word cloud
273 votes

Poll settings

What is your go-to data visualization tool? (E.g., spreadsheet, ggplot, Power BI)

Review answers

- spreadsheet 79
- Excel 73
- Power BI 46
- Excel spreadsheet 17
- plotly/ggplot2 12
- Tableau 11
- ggplot 9
- PowerPoint 7



Introduction

Not a new topic…

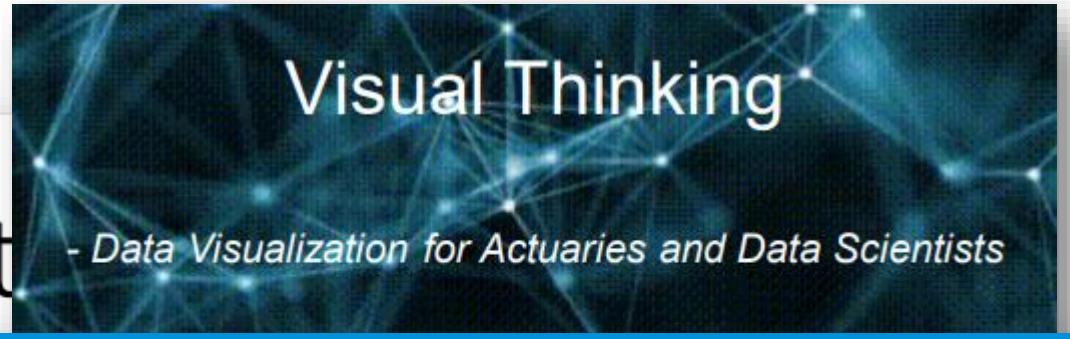
Death by Chartjunk? Graphical Excellence In Insurance

CAS 2012 Annual Meeting

Data Visualizat

The importance of princi

Keith Quigley, A



Picture This:
Using Data Visualization

A graph is worth

The effective use of visu

CLRS 2022 St. Louis

Jamie Mackay

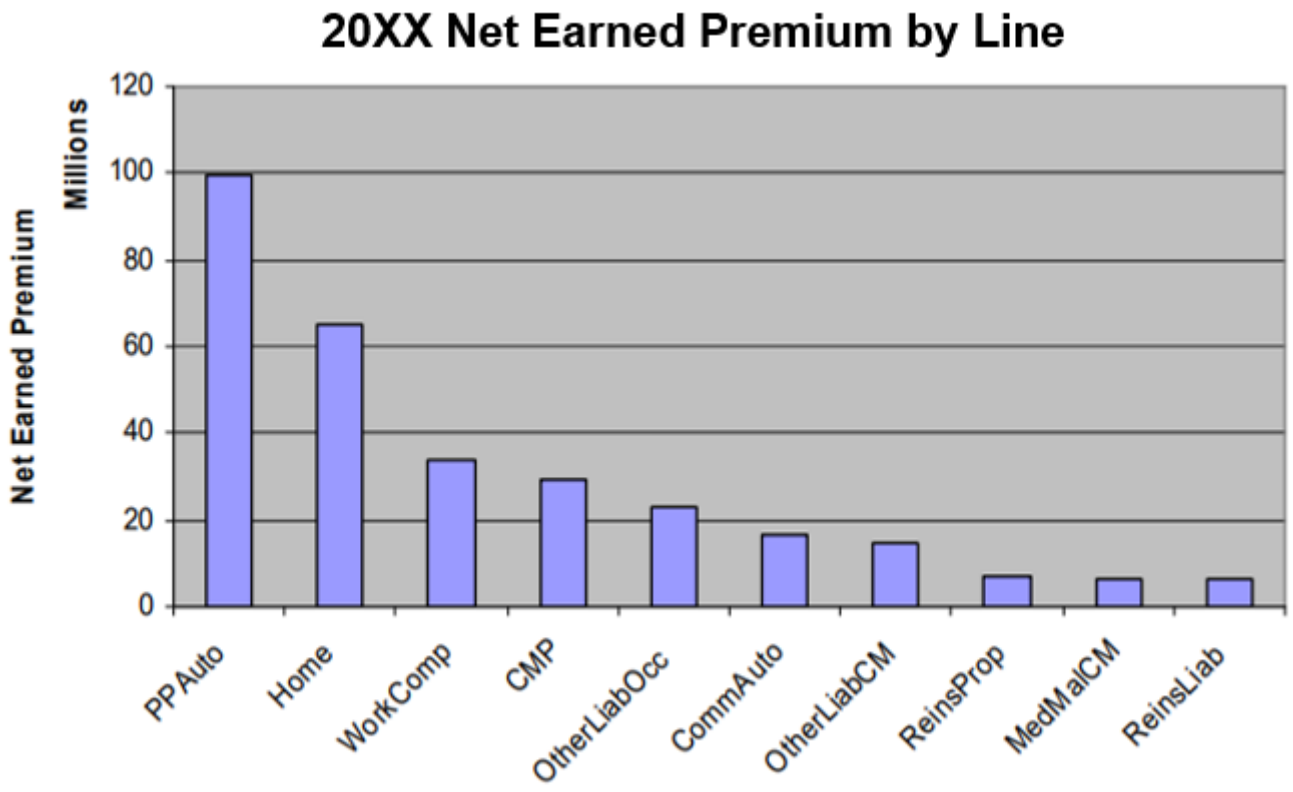
Introduction to Data Visualization

John Deacon, Annie Fan, Brian Fannin, Jennifer Levine, Keith Quigley, Patrick Yu

Abstract: This paper summarizes some of the literature on the topic of basic data visualization techniques. We emphasize the importance of knowing the audience, and focusing on what message is intended to be sent. We provide visual examples of graph types and describe when to use the different types for different situations. We identify several decluttering and accentuating techniques and we share some of the basic research on how the human eye and brain work to interpret visual information. We provide a before-and-after example of the basic data visualization techniques, to show how much improvement can be achieved in delivering the intended message.

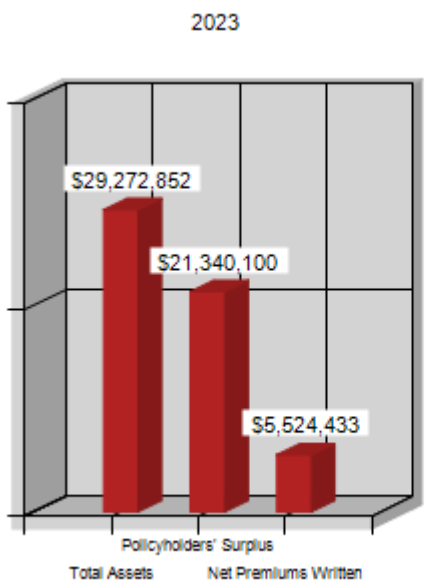
keywords: data visualization, communication, gestalt principles

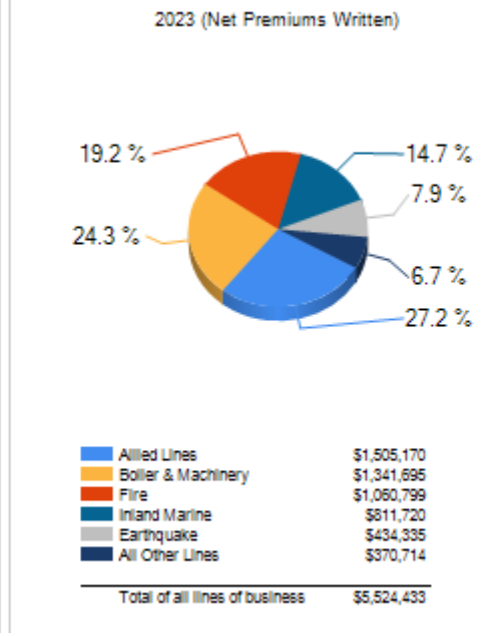
But still seeing this?

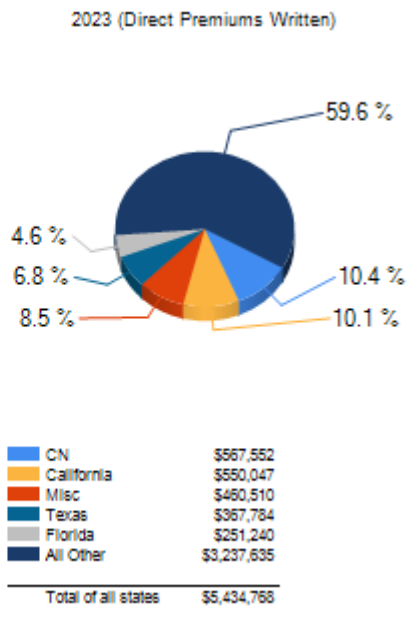


Annual Statement Data

Or this?

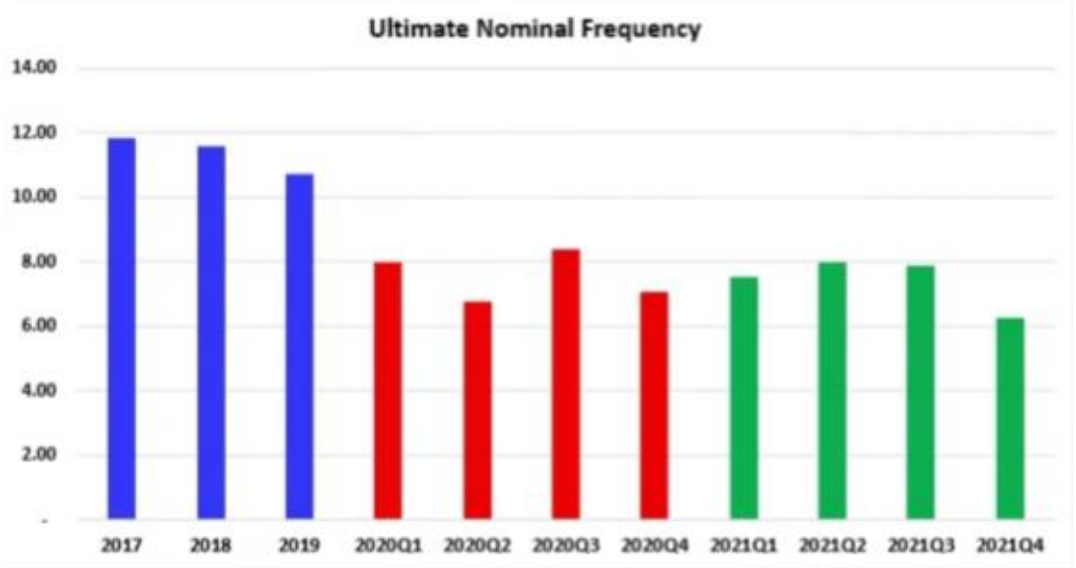


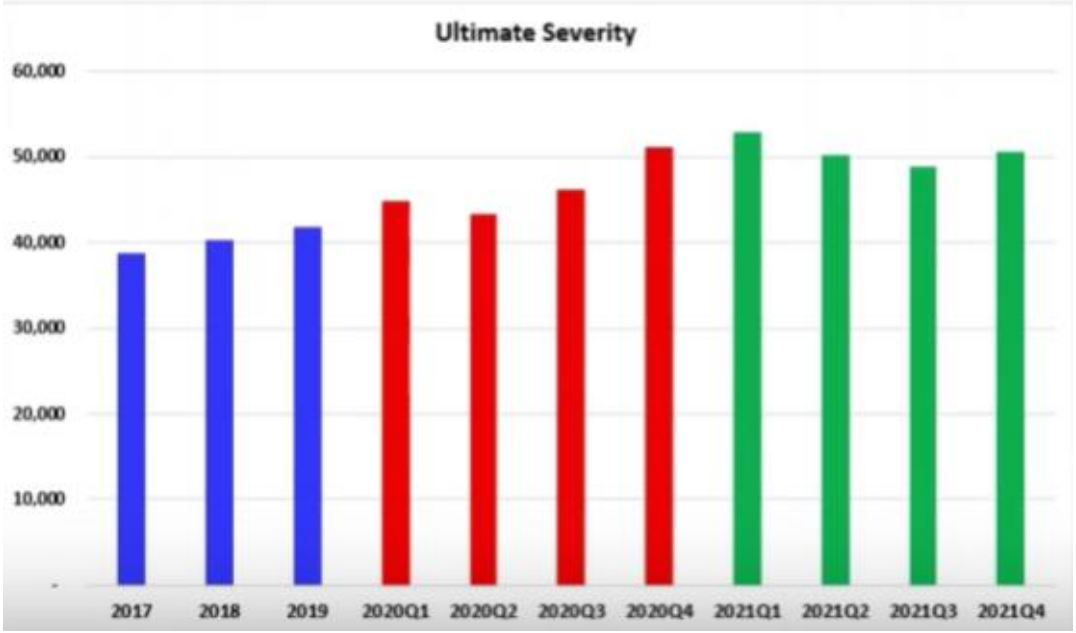




An unnamed rating agency

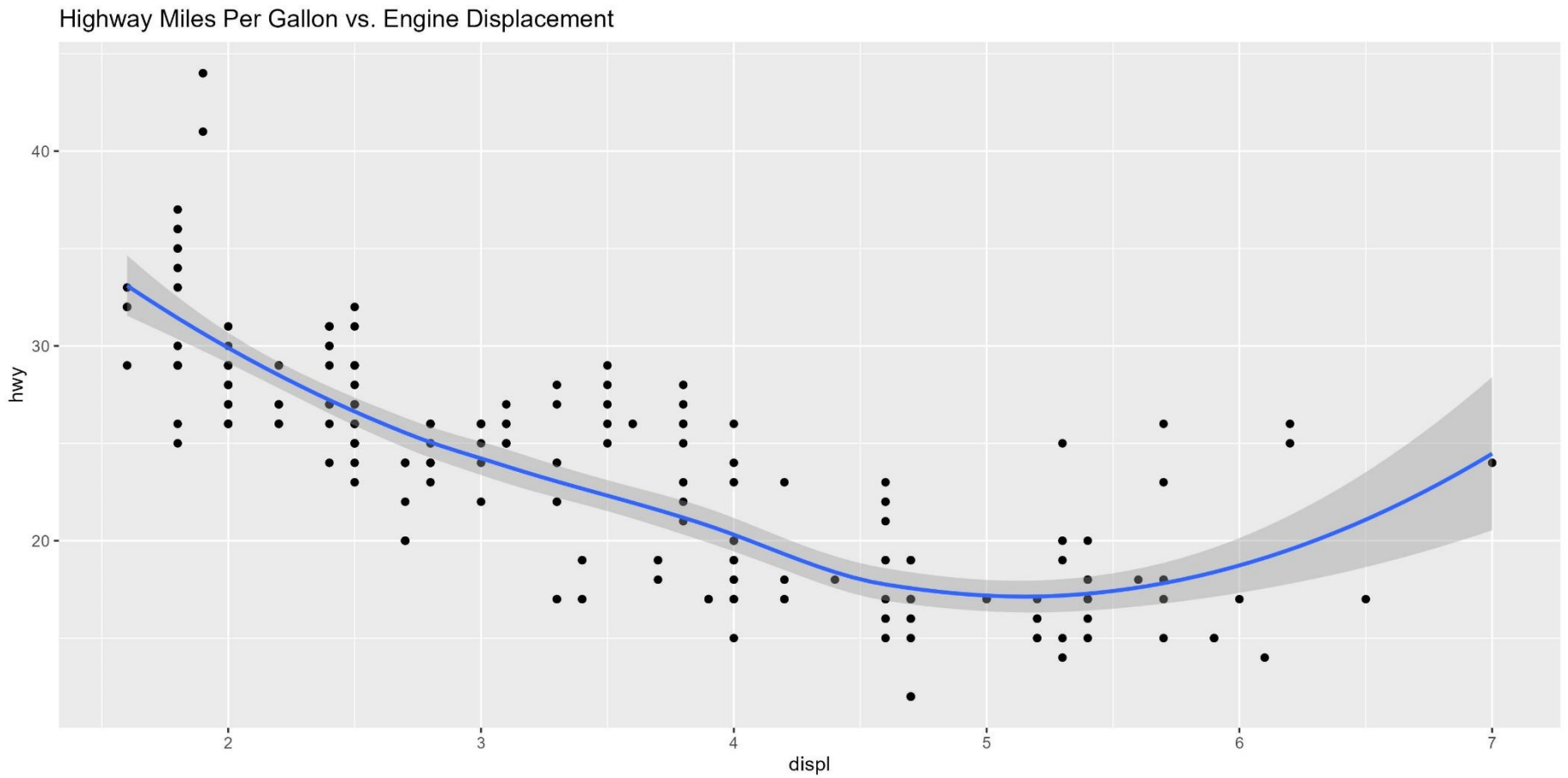
Or this?





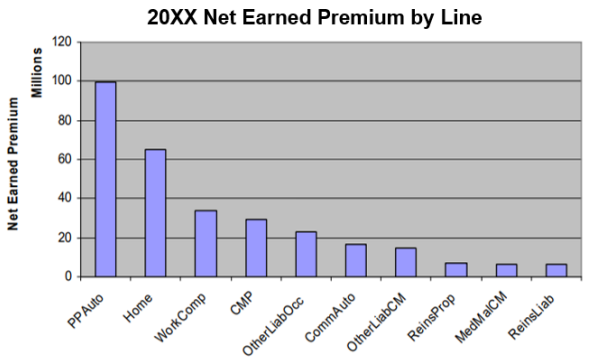
An unnamed industry benchmark provider

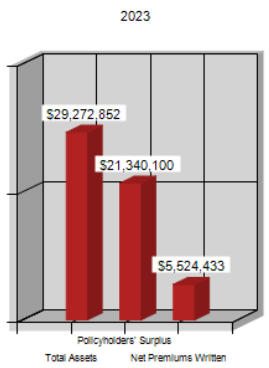
What about this?

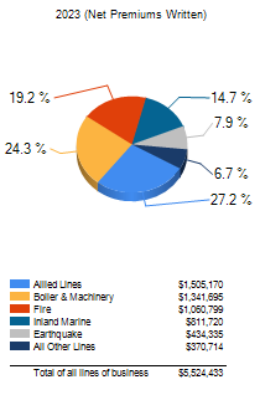


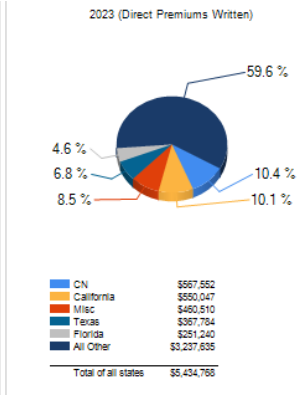
https://ggplot2.tidyverse.org/reference/mpg.html

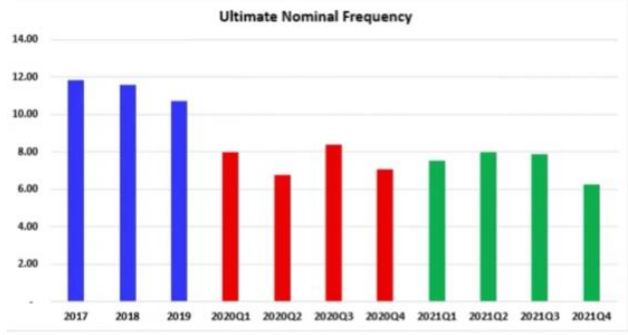
How can we do better?

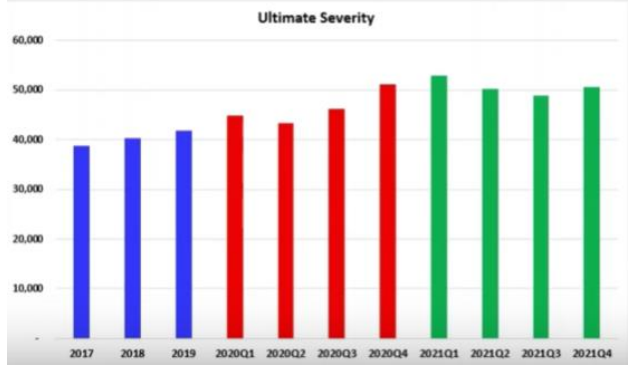


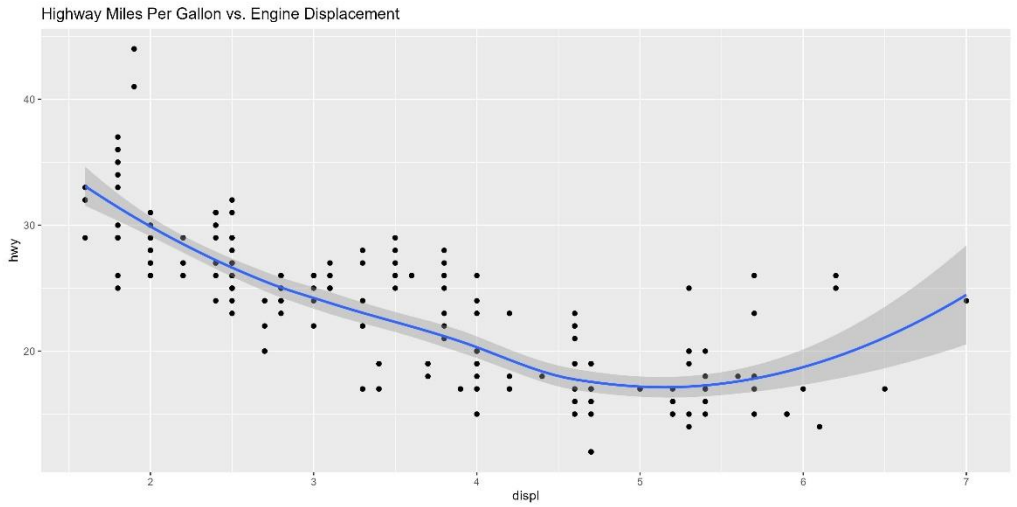






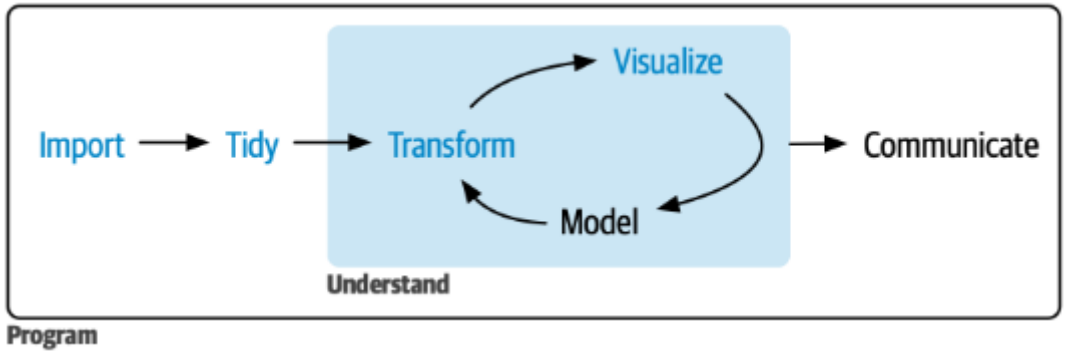






Why should we care?

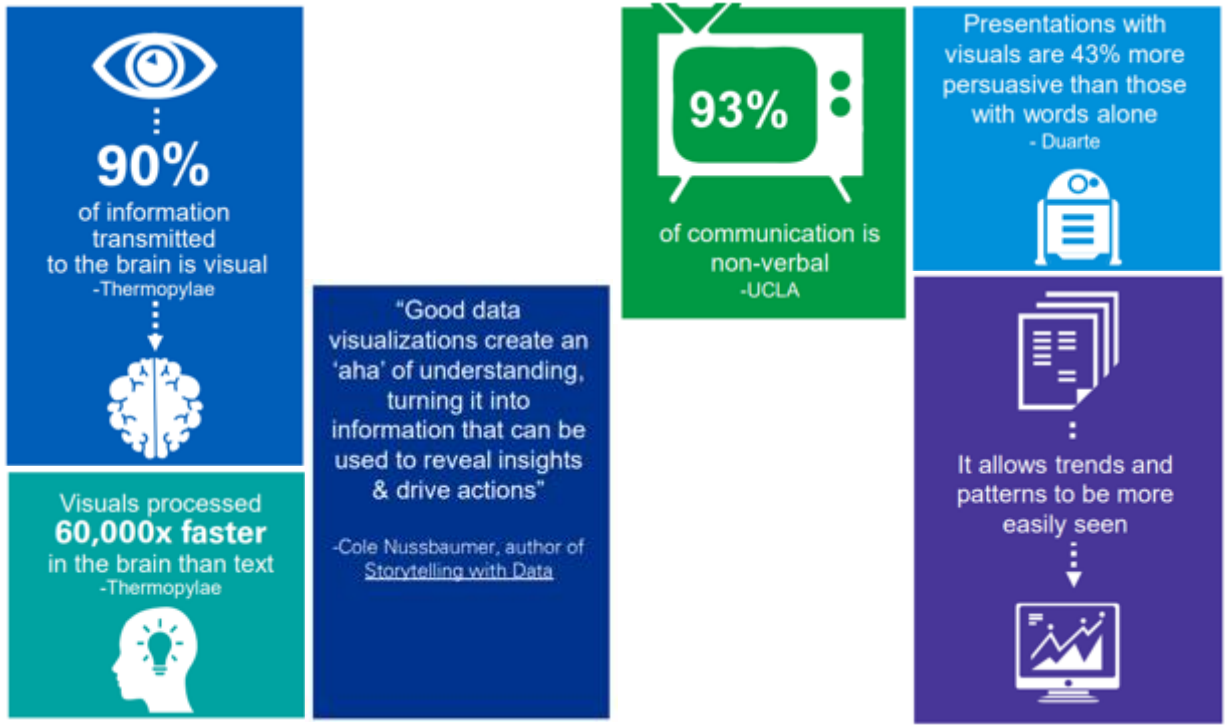
- Because “visualize” is an integral part of the actuarial/data modeling process, and arguably it’s critical to every step of the process



R for Data Science (2e) - Whole game (hadley.nz)

Why should we care?

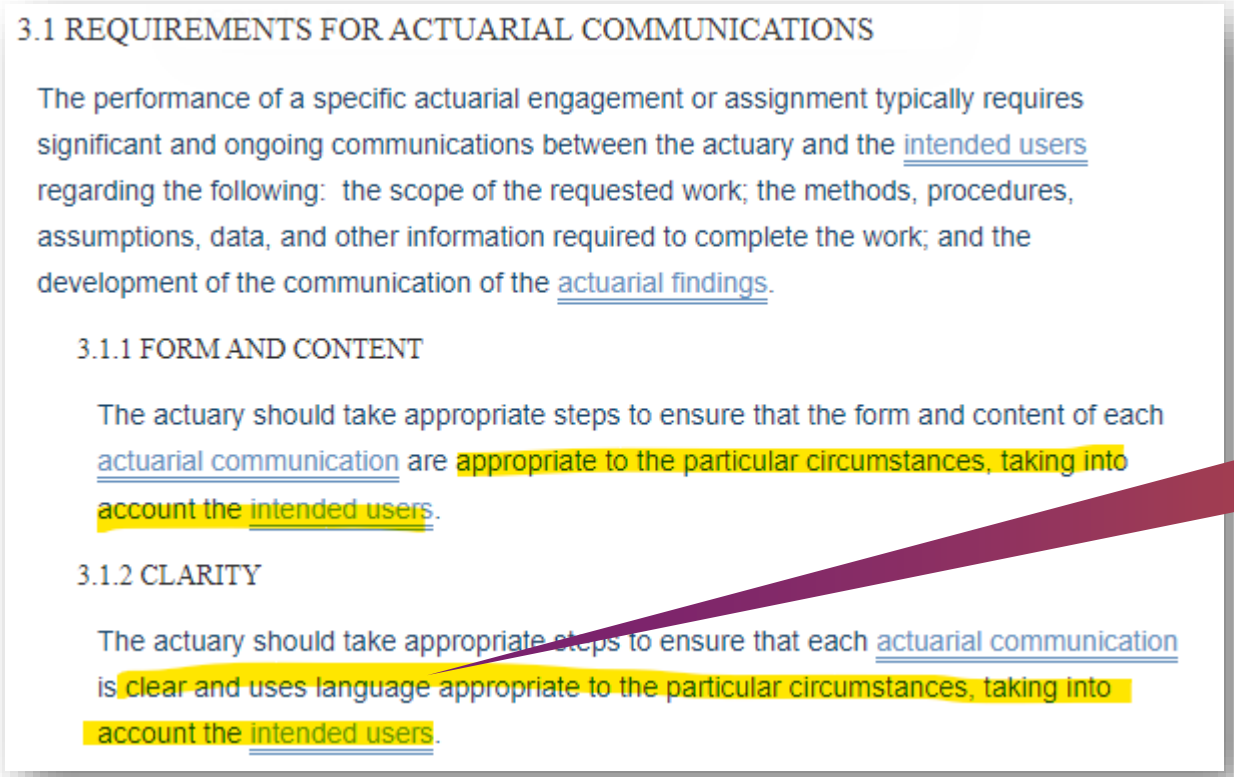
- Because **good** visualization **could** enhance effectiveness of communication



cs17-picturethis_usingdatavisualizationtodriveresults-fileid-255085.pdf

Why should we care?

- Because we are actuaries and ASOPs says…



3.1 REQUIREMENTS FOR ACTUARIAL COMMUNICATIONS

The performance of a specific actuarial engagement or assignment typically requires significant and ongoing communications between the actuary and the intended users regarding the following: the scope of the requested work; the methods, procedures, assumptions, data, and other information required to complete the work; and the development of the communication of the actuarial findings.

3.1.1 FORM AND CONTENT

The actuary should take appropriate steps to ensure that the form and content of each actuarial communication are appropriate to the particular circumstances, taking into account the intended users.

3.1.2 CLARITY

The actuary should take appropriate steps to ensure that each actuarial communication is clear and uses language appropriate to the particular circumstances, taking into account the intended users.

Visualization may be the most appropriate “language” in certain circumstances for the intended users

Actuarial Communications - Actuarial Standards BoardActuarial Standards Board

Not a new topic, so what's new(-ish)?

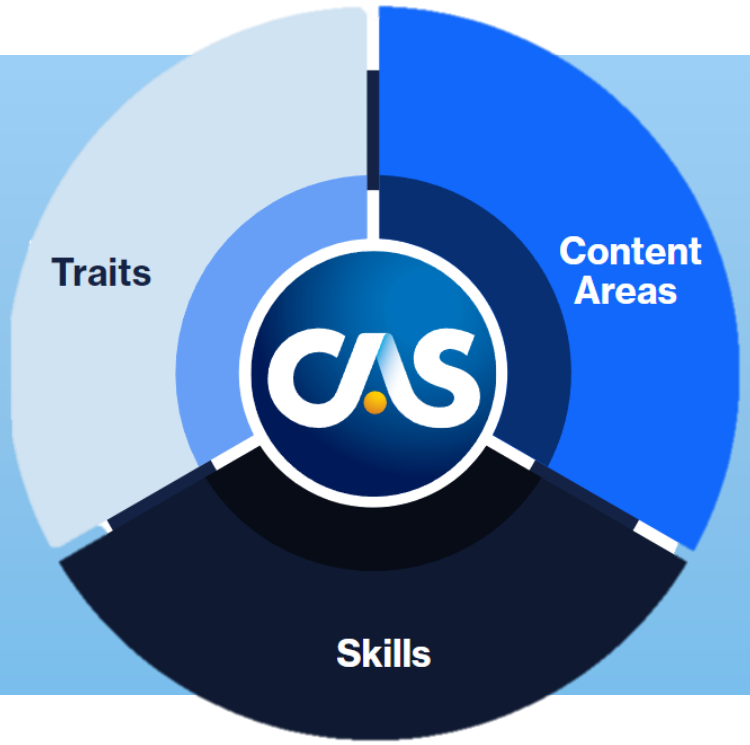
- Data visualization wasn't part of the exam syllabus at all until very recently through the new online course

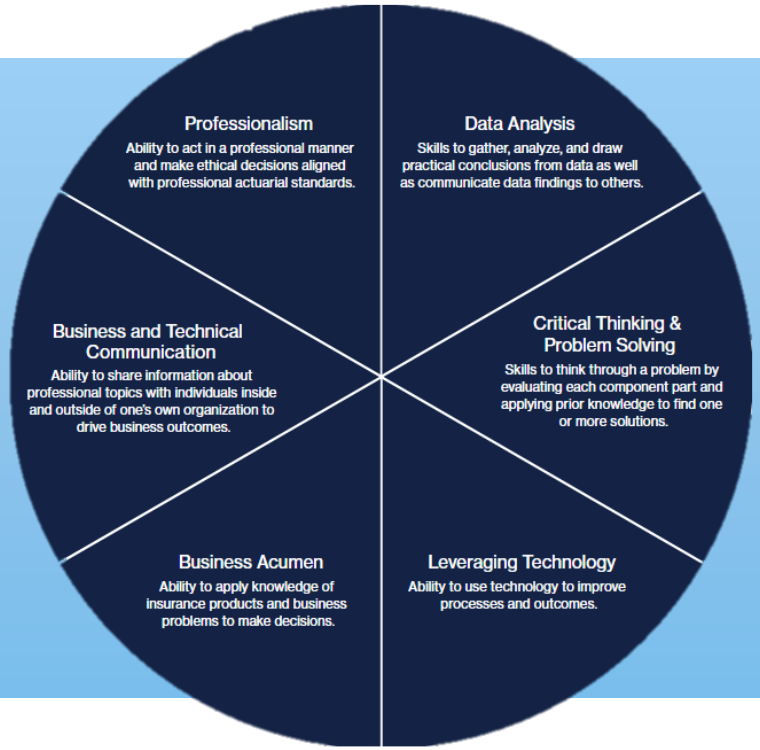
5. Analyzing Data With Visualizations
    a. Planning an Effective Data Exploration
    b. Data Exploration Fundamentals
    c. Fundamentals of Exploratory Data Visualizations
    d. Creating Plots

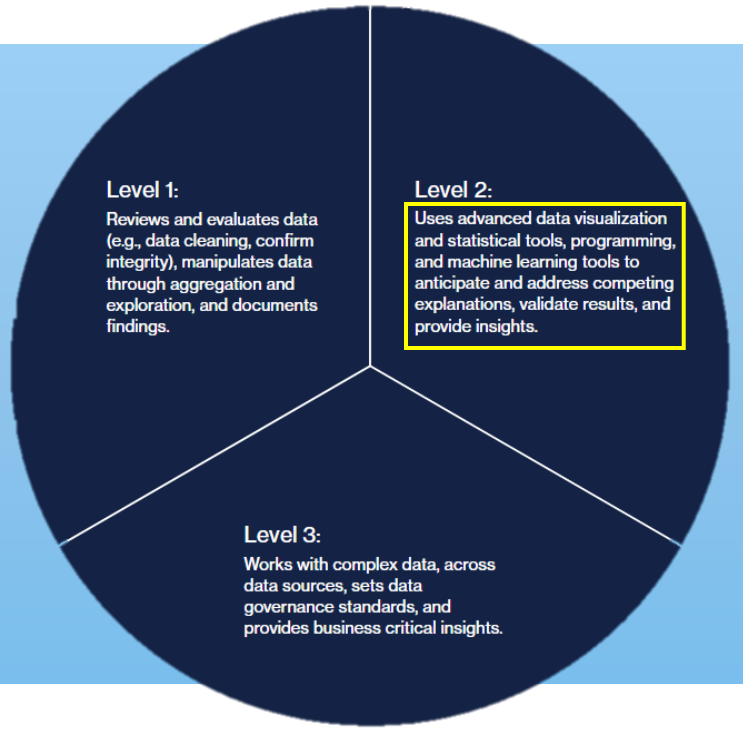
- Growing popularity of new and free tools (i.e., R, Python) that come with much more robust visualization libraries than Excel

CAS DISC DA Syllabus (casact.org)

CAS Capability Model







CAS Capability Model | Casualty Actuarial Society (casact.org)

Main learning objectives

1. Understand basic theories and principles of data visualization and what constitutes a “Good Chart”
2. Apply the learning from this session to create desired visuals. We will use *ggplot* for practical demonstration, but the same principles can be extended to other platforms
3. Evaluate the appropriateness of a visualization in a given context and propose ways to improve it

What this session is not

- A tutorial/workshop for *ggplot*, *Power BI*, etc.
  - Choose your favorite tool, though we will be using *ggplot* for demonstration of the principles discussed herein
- A sales pitch on why and how your organization should set up a more efficient dashboard system
  - Consultants are more than happy to take your money and help you with that

Theories and Principles

Tufte's Graphical Excellence

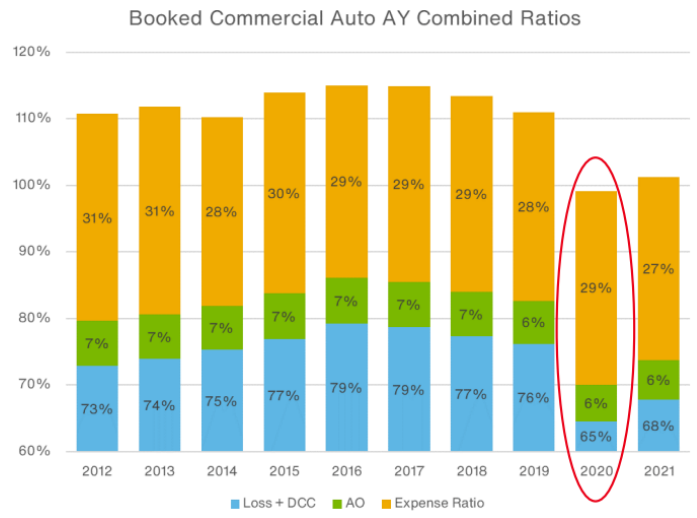
- Graphical excellence is that which gives to the viewer the **greatest number of ideas in the shortest time** with the least ink **in the smallest space**
  - The famous "data-ink ratio" stems from this
  - Higher data "density," if you will
- Graphical excellence requires telling the **truth**
- …
- "…above all else, show the data"

Graphical/Visual Properties (gatech.edu)

Visuaries do not mislead

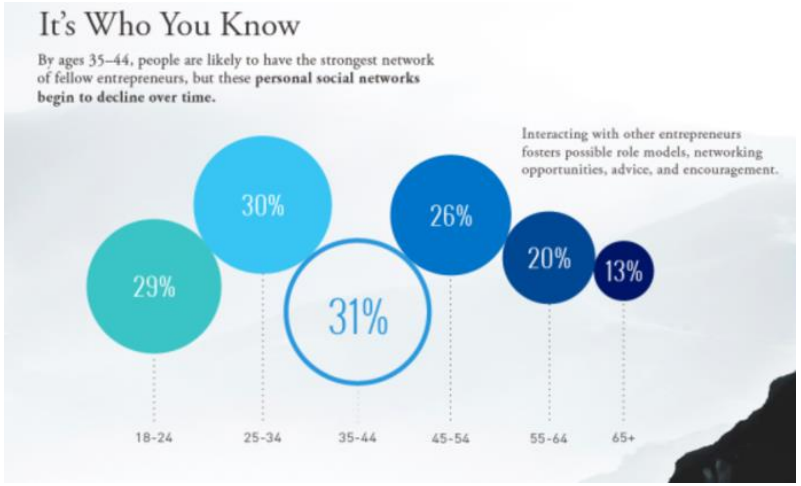
- “Graphical excellence requires telling the truth about the data.”
  - Visual Display of Quantitative Information pg. 51

Obvious



PowerPoint Presentation (casact.org)

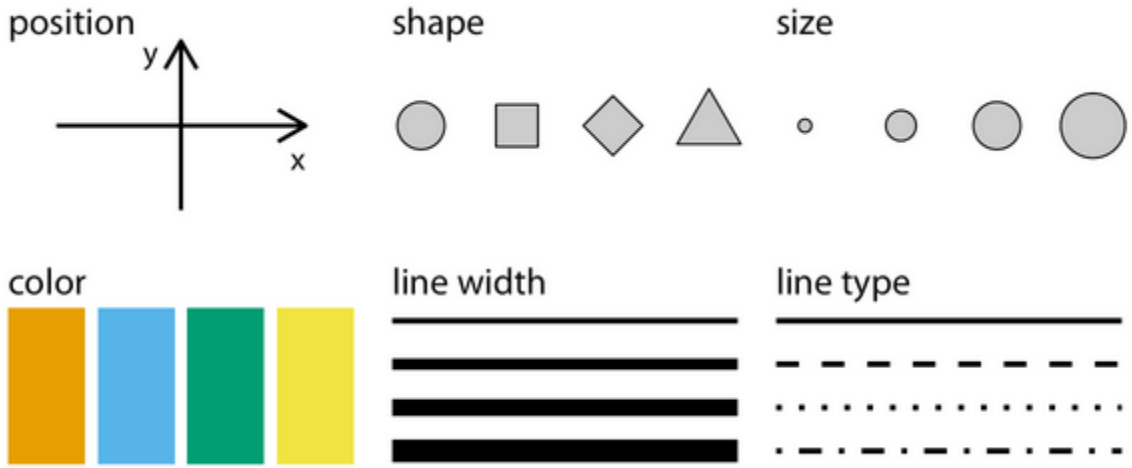
Subtle



Tools - Proportional Ink (callingbullshit.org)

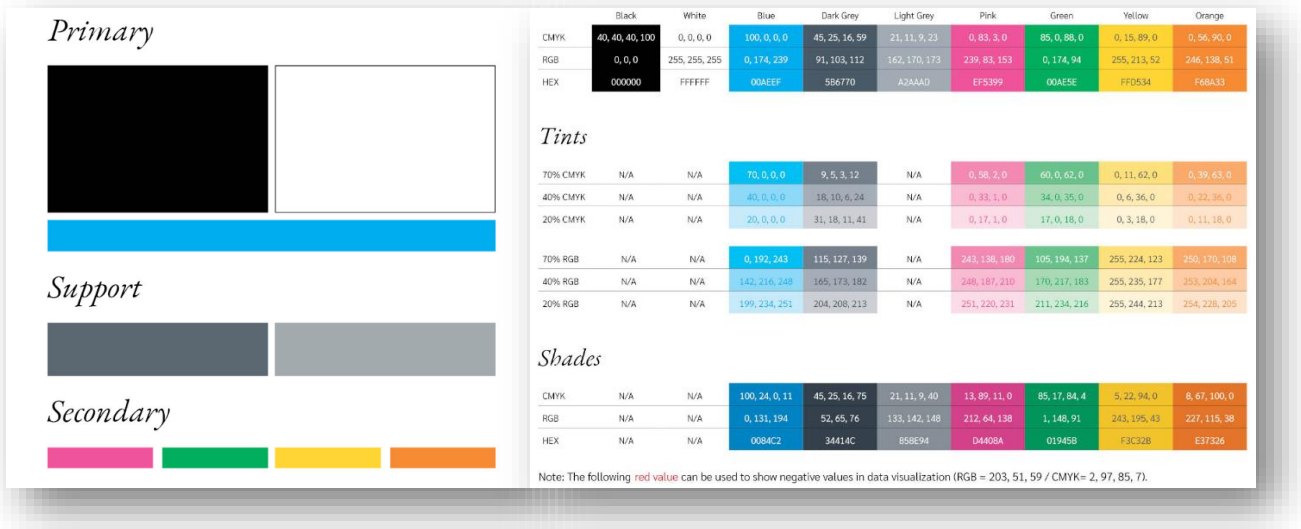
Visuaries employ intentional aesthetics

- "Aesthetics"
  - *aes()* in *ggplot*
  - Combine colors with other aesthetics to ensure plot remains interpretable when printed black-and-white and for colorblind audience
  - There are also colorblind-proof palettes available, e.g., Okabe-Ito



Color Considerations

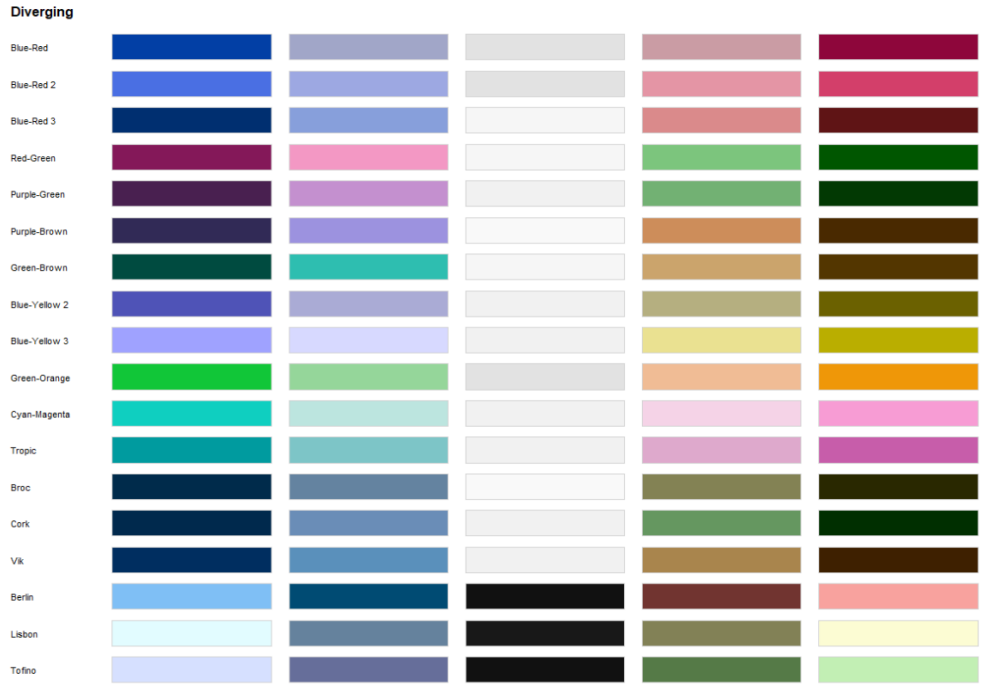
- Usually dictated by your companies' themes
  - If so, be on-brand!



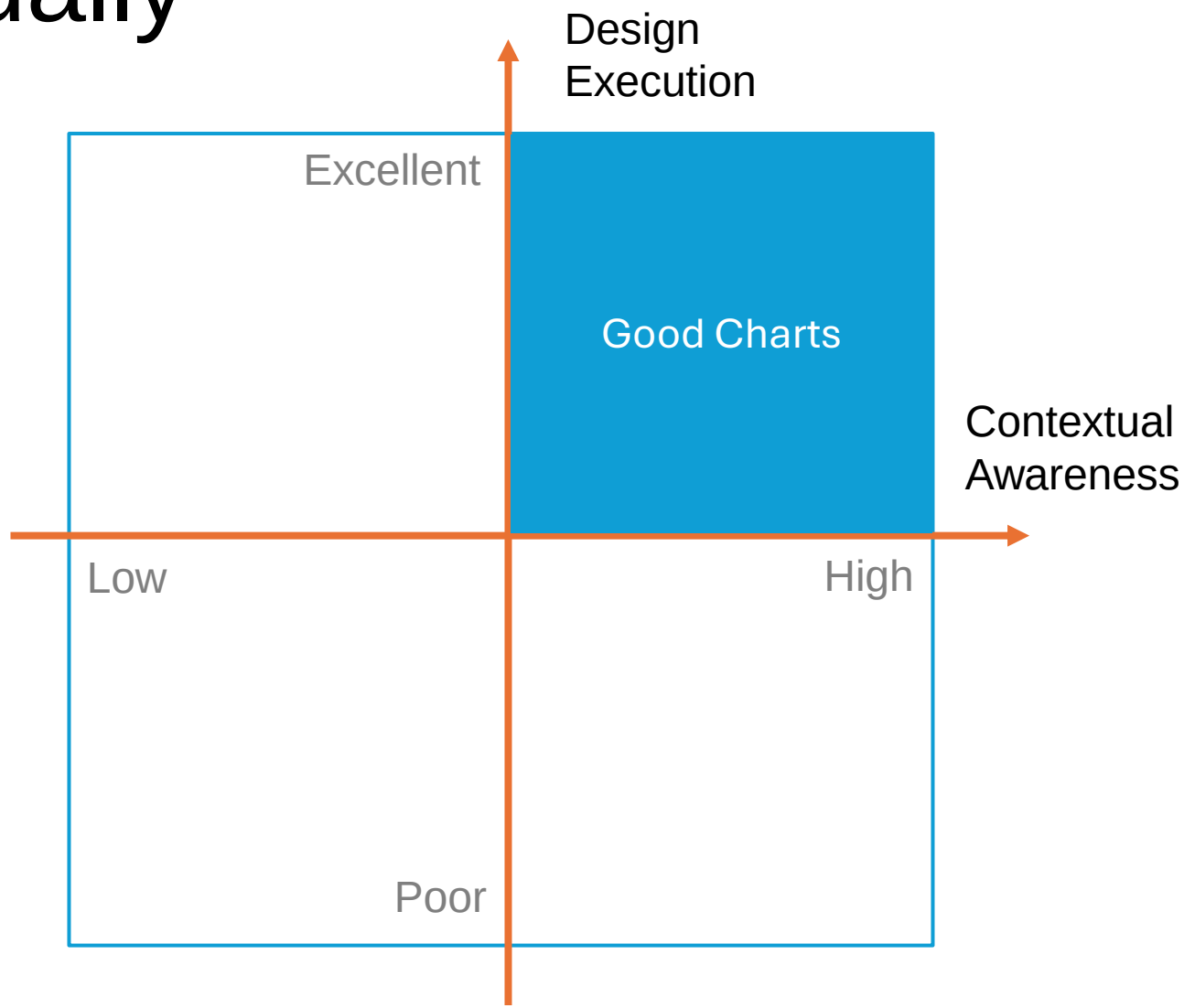
  - If not, understand the use cases:
    - Categories (qualitative), numeric values (sequential/diverging), highlight

Color Considerations

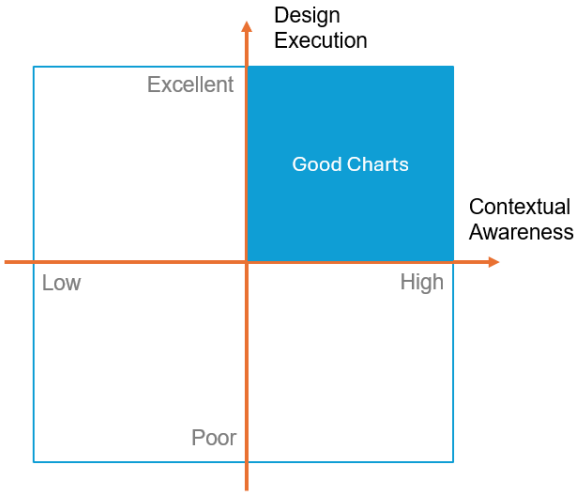
```
colorspace::hcl_palettes(type = "Sequential", plot = TRUE)
colorspace::hcl_palettes(type = "Diverging", plot = TRUE)
colorspace::hcl_palettes(type = "Qualitative", plot = TRUE)
```



Visuaries design excellently and contextually



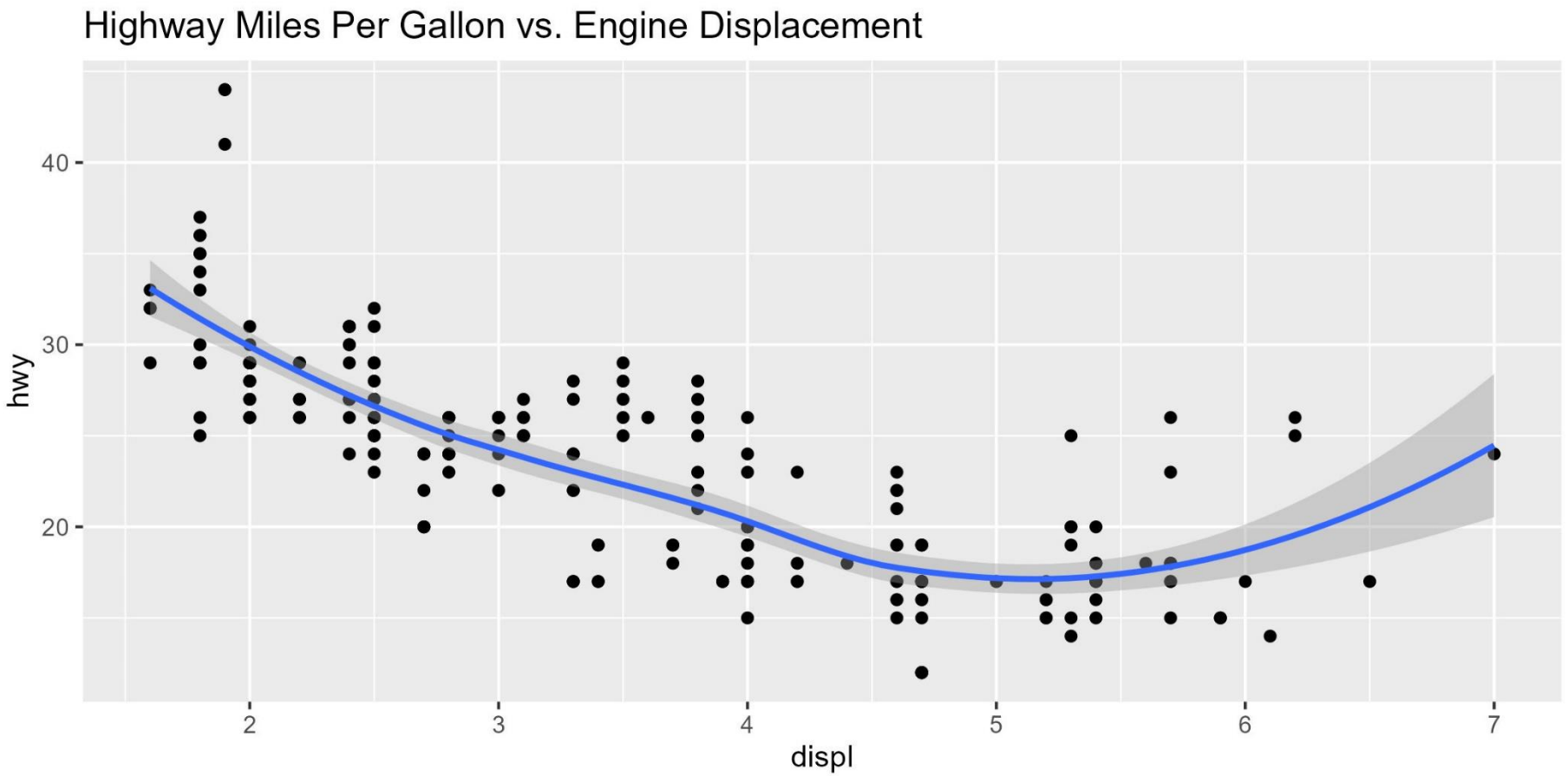
Good Charts (scottberinato.com)



Design Execution

Reproducible R script available on GitHub

Original

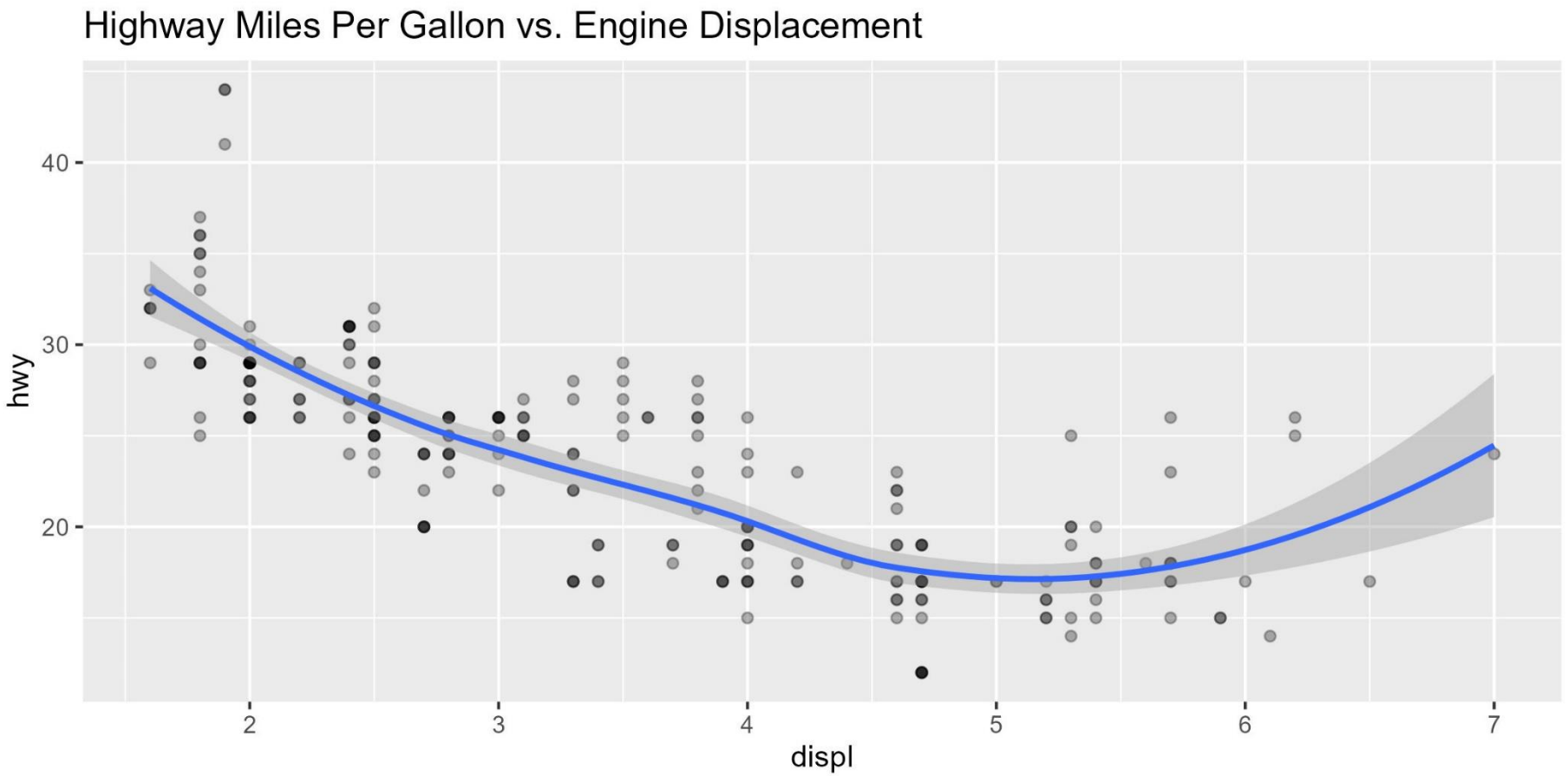


> How would you improve this plot?

Step 1: Look at your data!

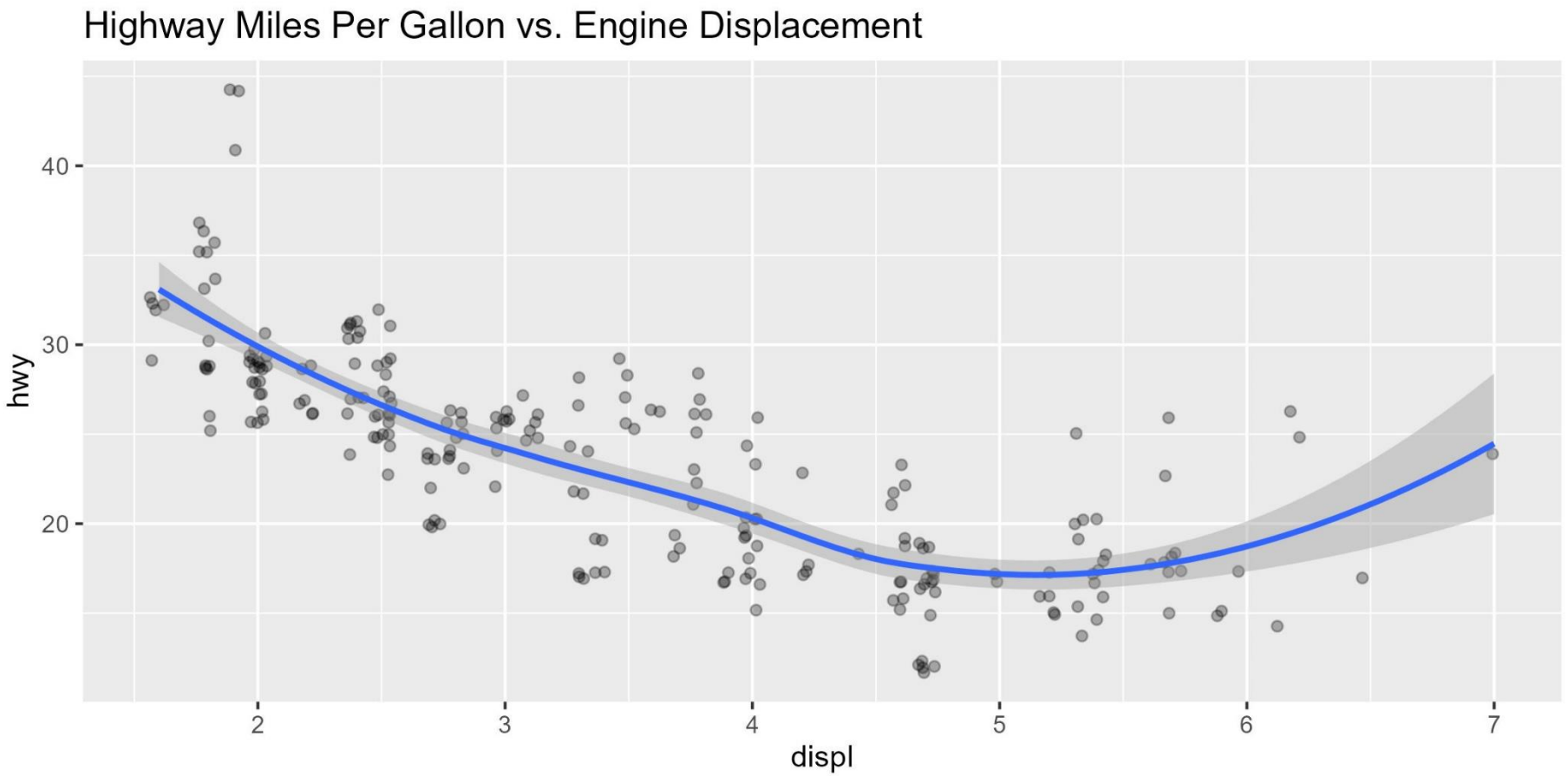
```
> mpg
# A tibble: 234 x 11
   manufacturer model      displ  year   cyl trans      drv     cty   hwy fl    class
   <chr>        <chr>      <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
 1 audi         a4           1.8  1999     4 auto(l5)   f        18    29 p     compact
 2 audi         a4           1.8  1999     4 manual(m5) f        21    29 p     compact
 3 audi         a4           2    2008     4 manual(m6) f        20    31 p     compact
 4 audi         a4           2    2008     4 auto(av)   f        21    30 p     compact
 5 audi         a4           2.8  1999     6 auto(l5)   f        16    26 p     compact
 6 audi         a4           2.8  1999     6 manual(m5) f        18    26 p     compact
 7 audi         a4           3.1  2008     6 auto(av)   f        18    27 p     compact
 8 audi         a4 quattro   1.8  1999     4 manual(m5) 4        18    26 p     compact
 9 audi         a4 quattro   1.8  1999     4 auto(l5)   4        16    25 p     compact
10 audi         a4 quattro   2    2008     4 manual(m6) 4        20    28 p     compact
# i 224 more rows
```

Intentional aesthetics - transparency

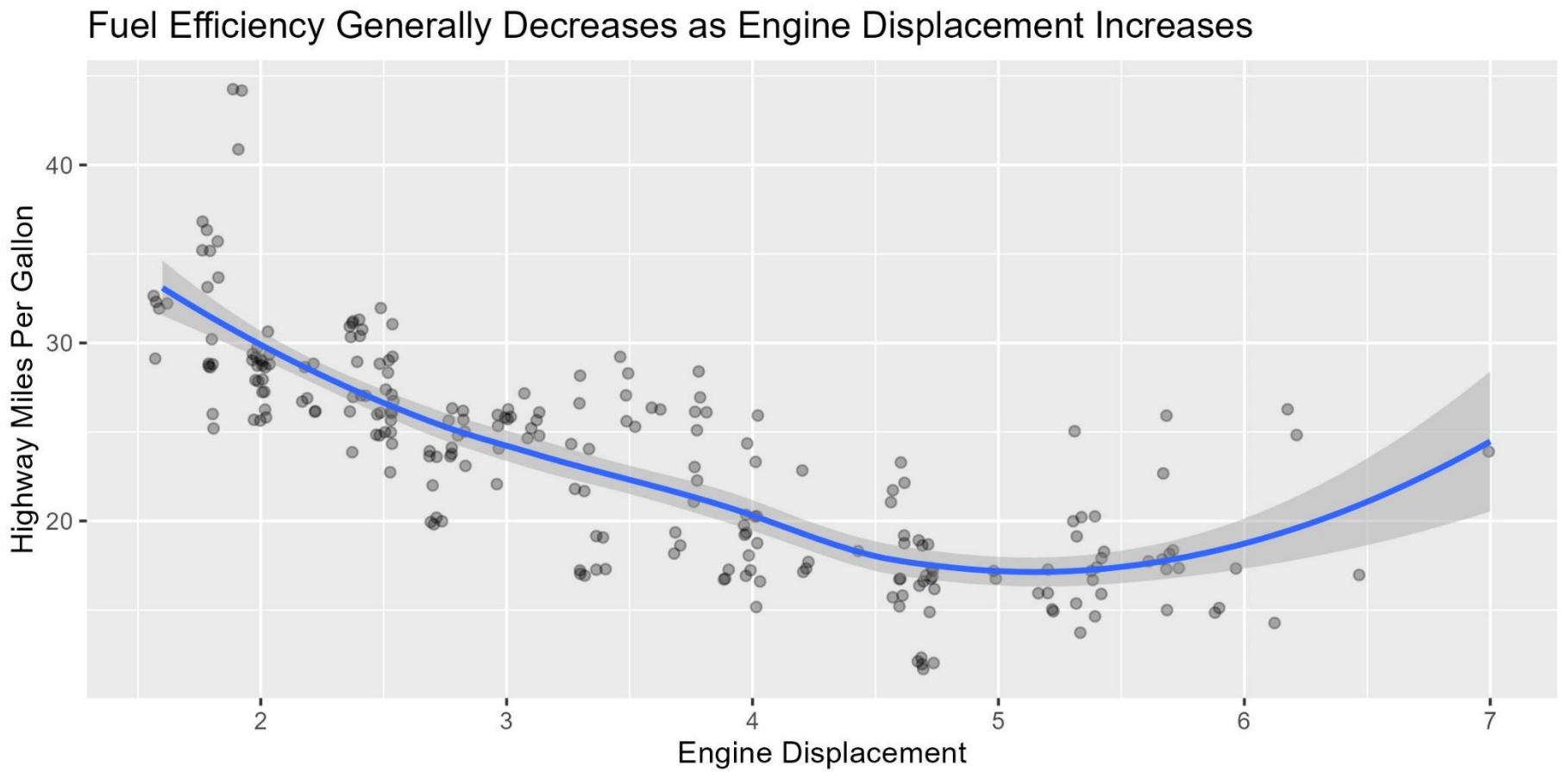


> "…above all else, show the data"

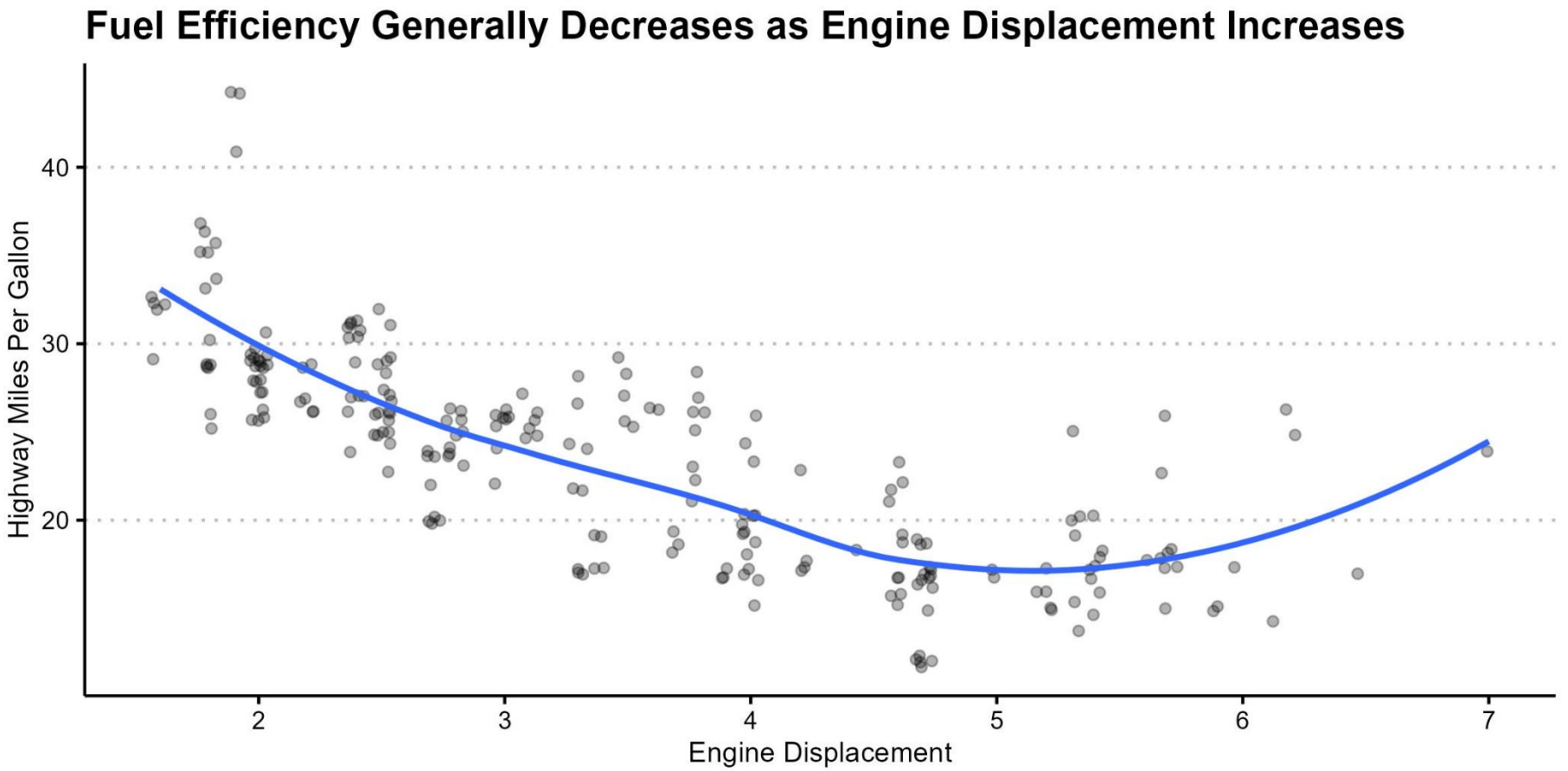
Intentional aesthetics - position



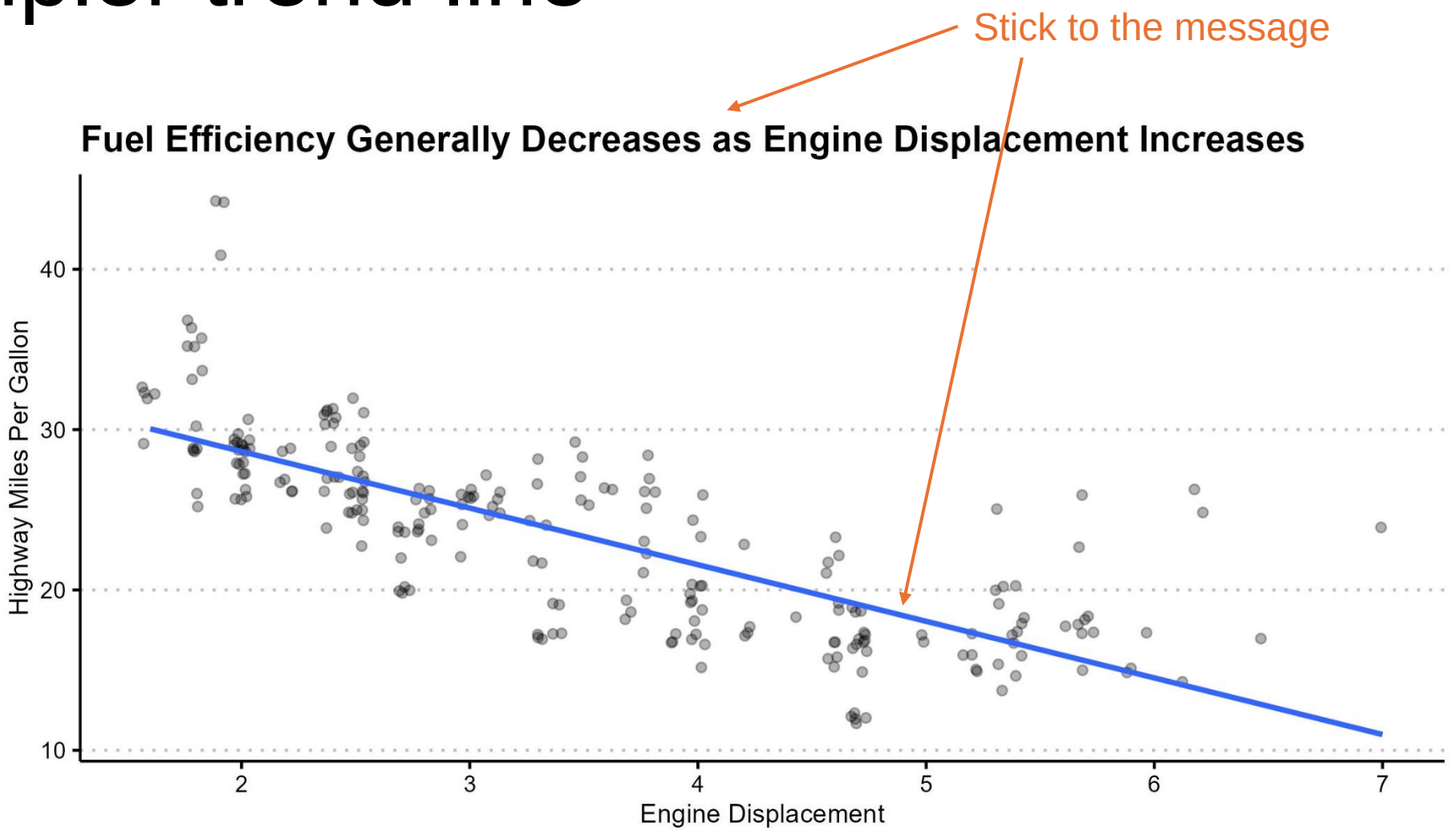
Descriptive labels and a storytelling title



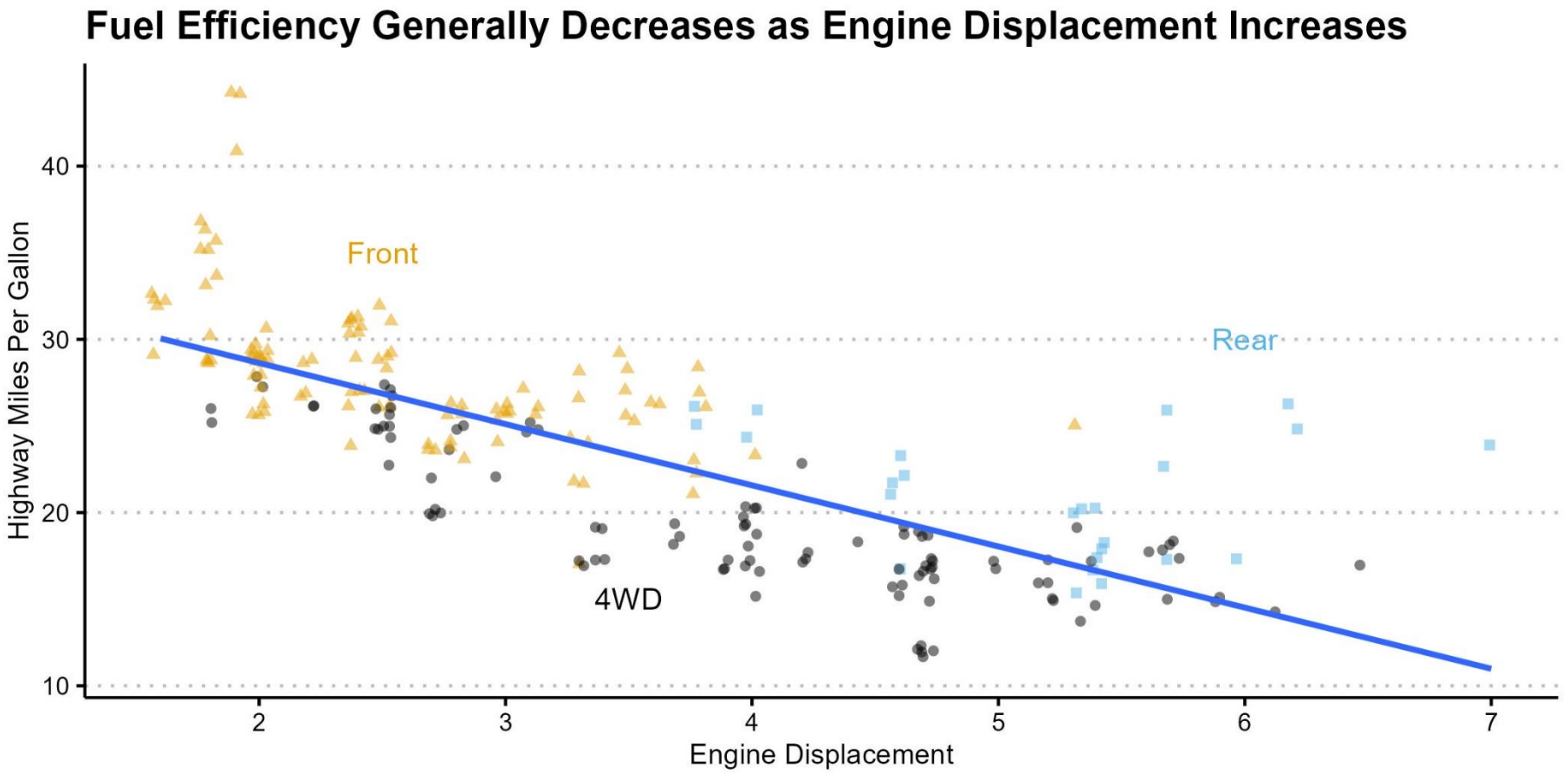
Remove distractions



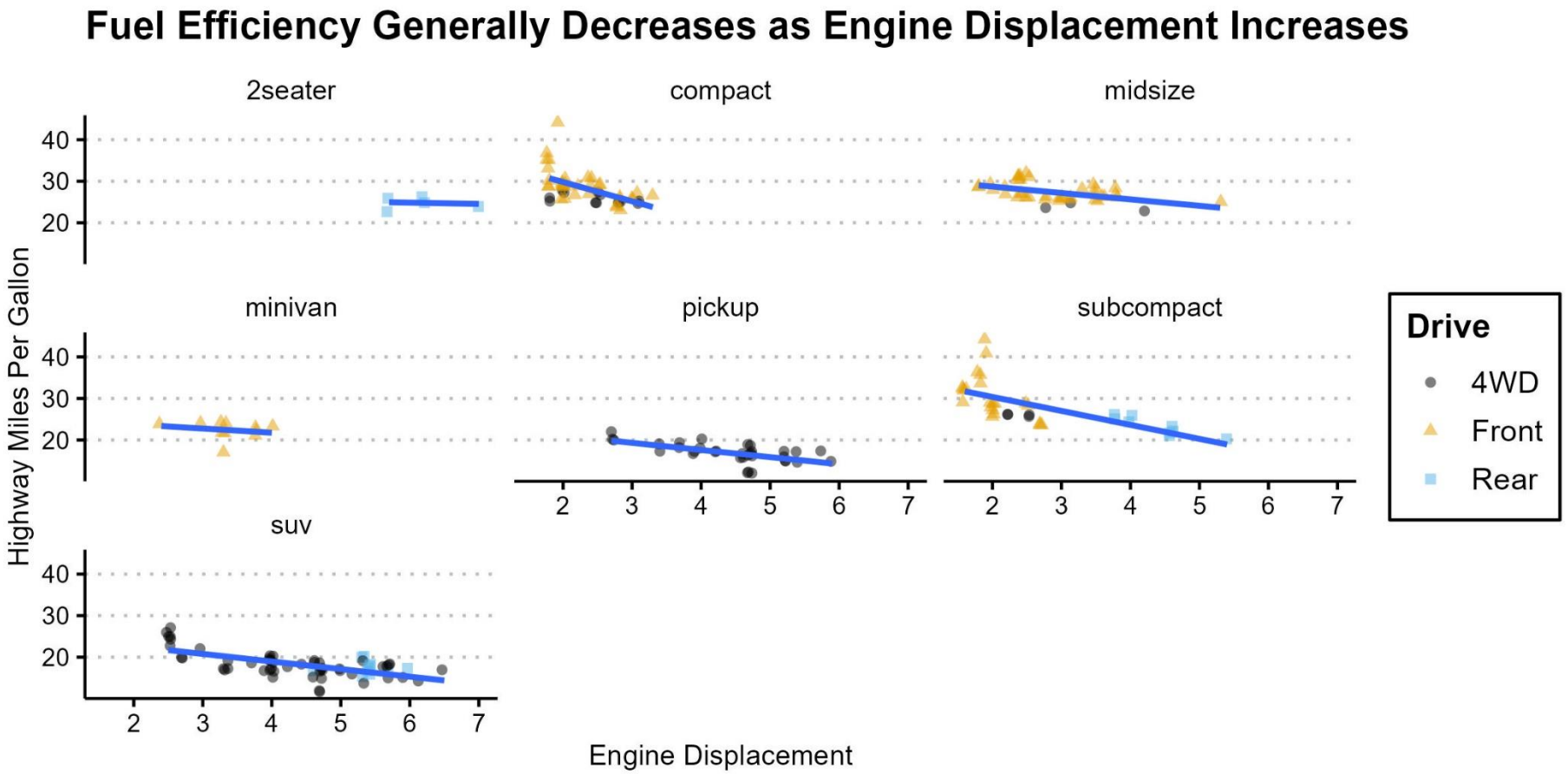
Simpler trend line



Increase data density – aesthetics

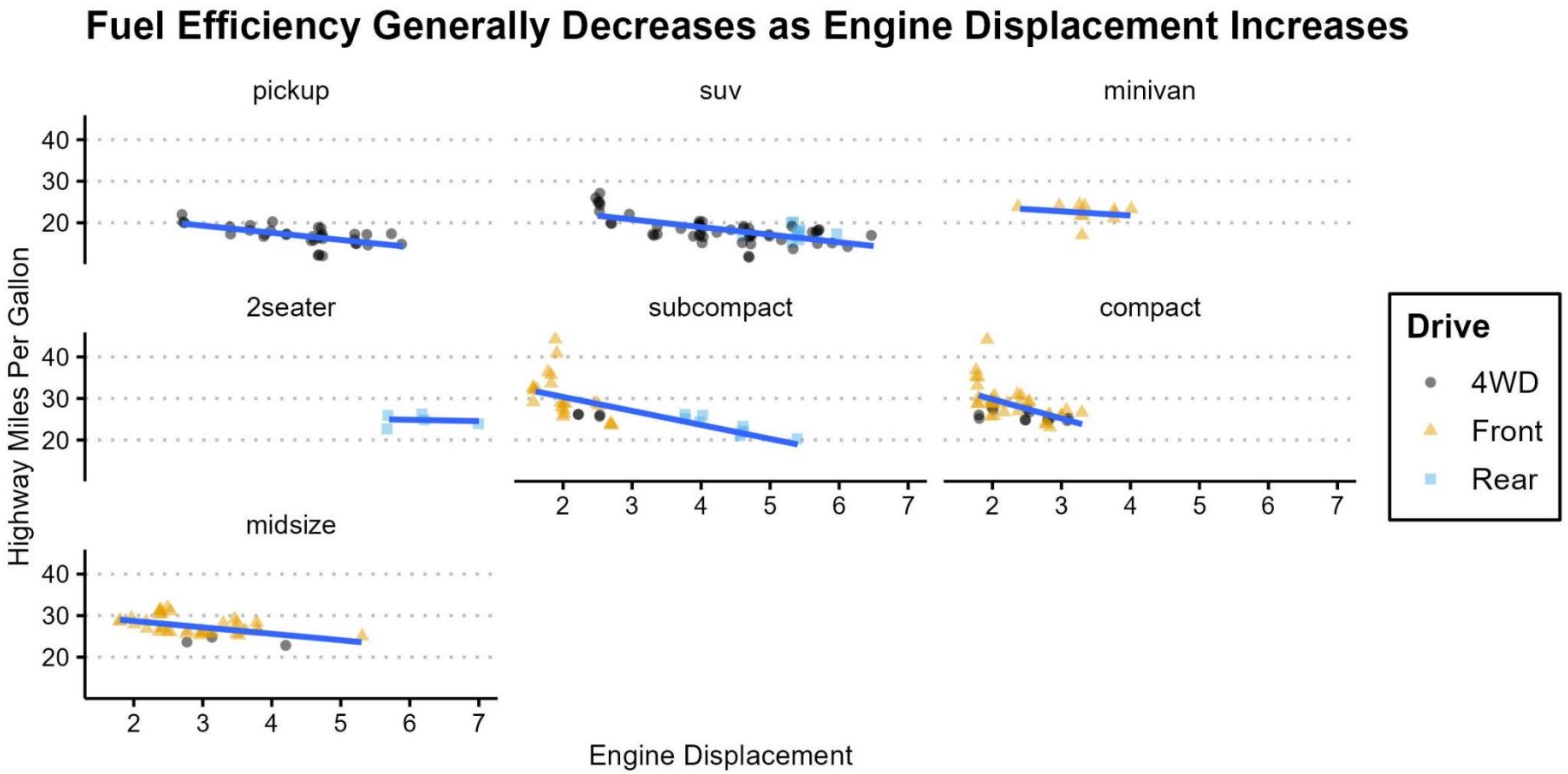


Increase data density – facets



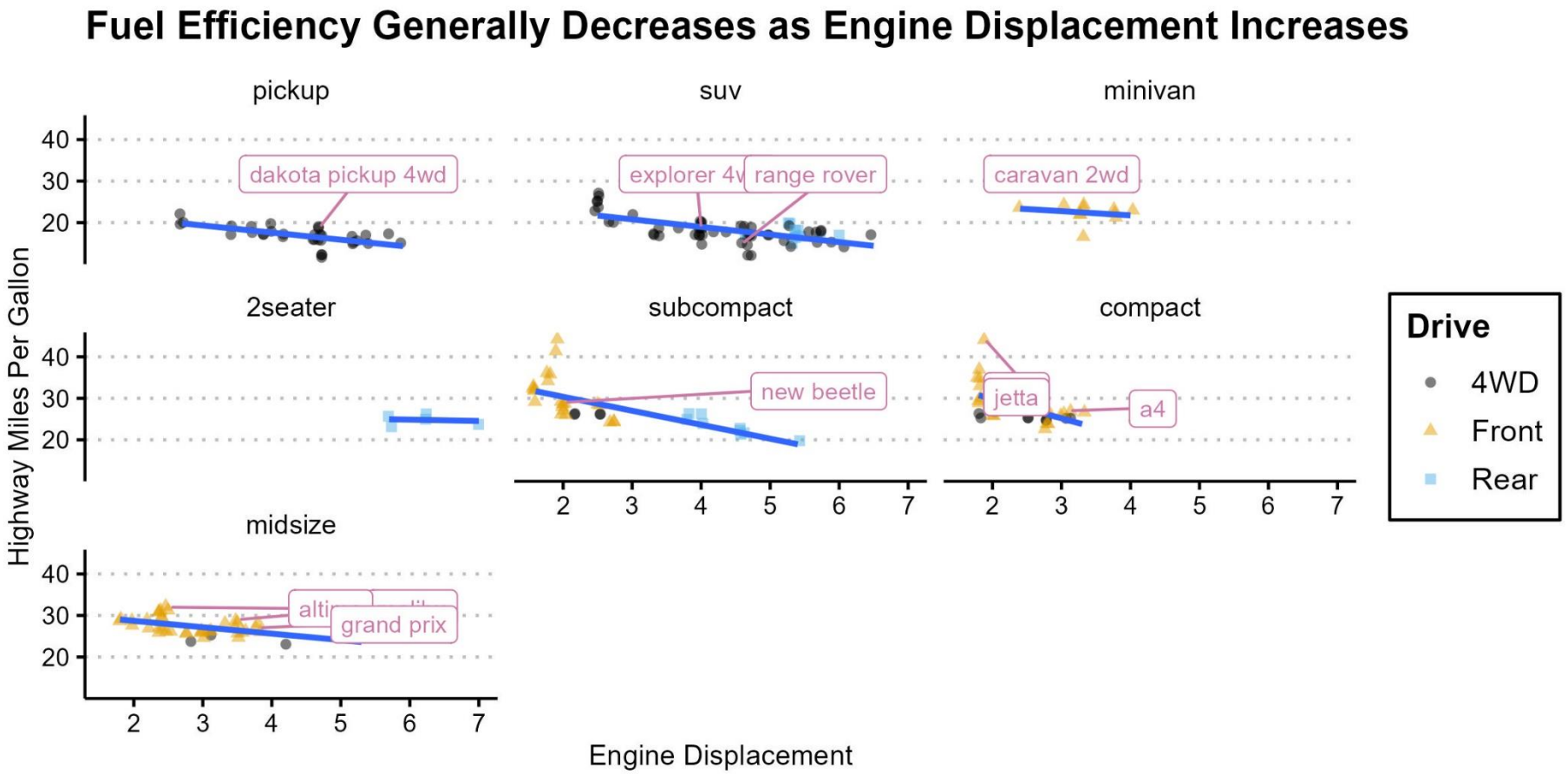
> Why should we not use alphabetical ordering for the panels?

Increase data density – order matters



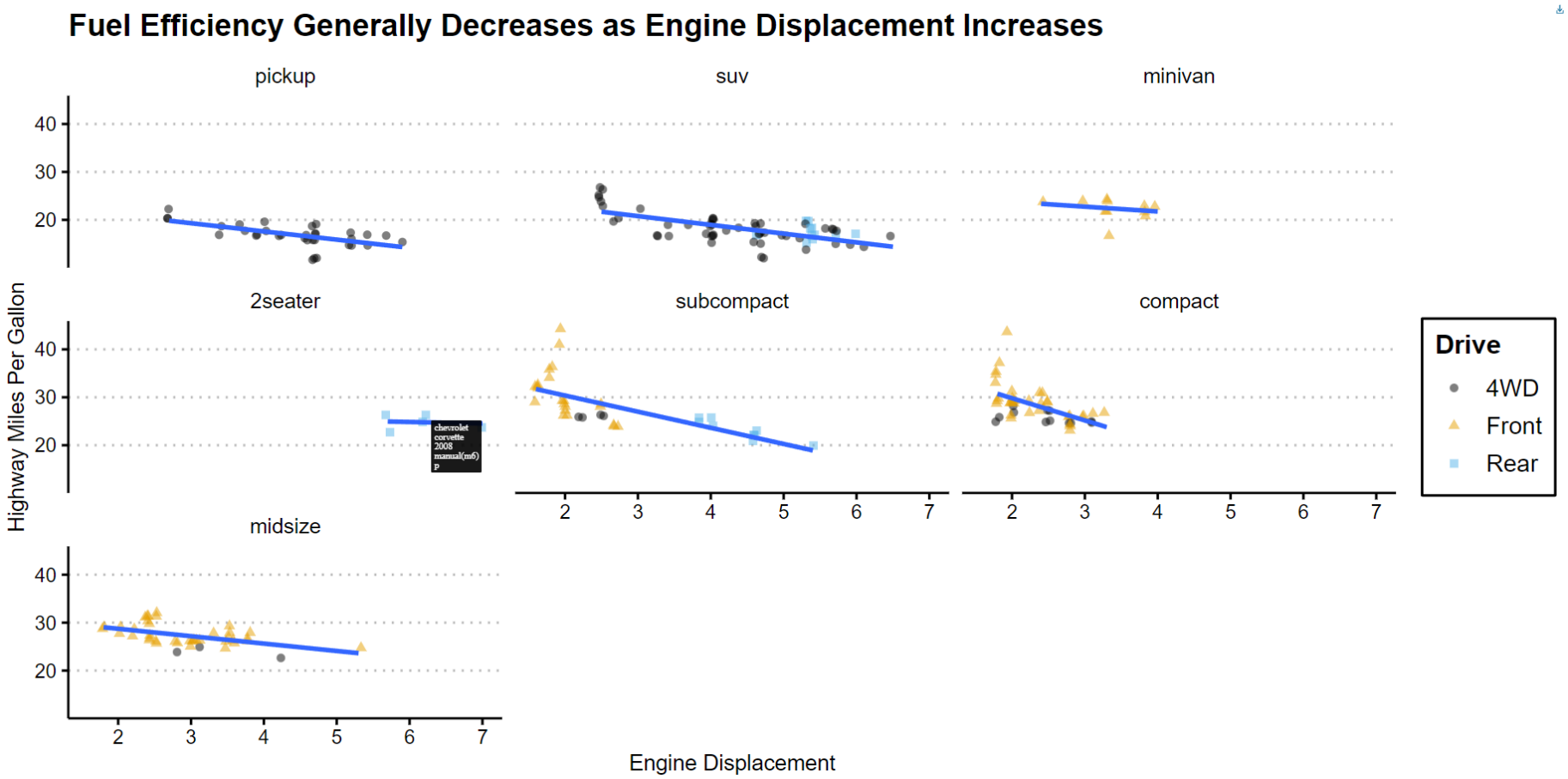
> Order panels by median Highway MPG of each class

Labels - static

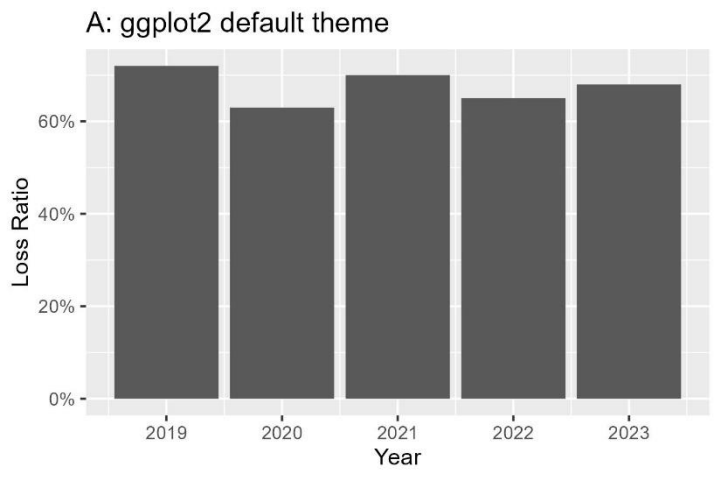


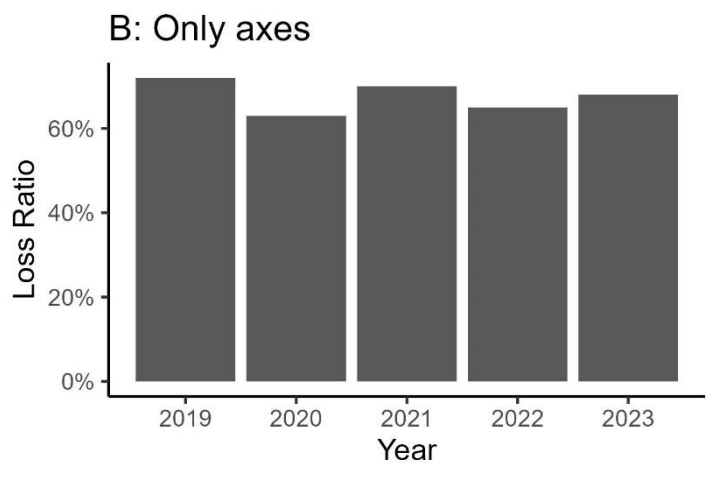
> Helpful but too busy. Anything else we could do?

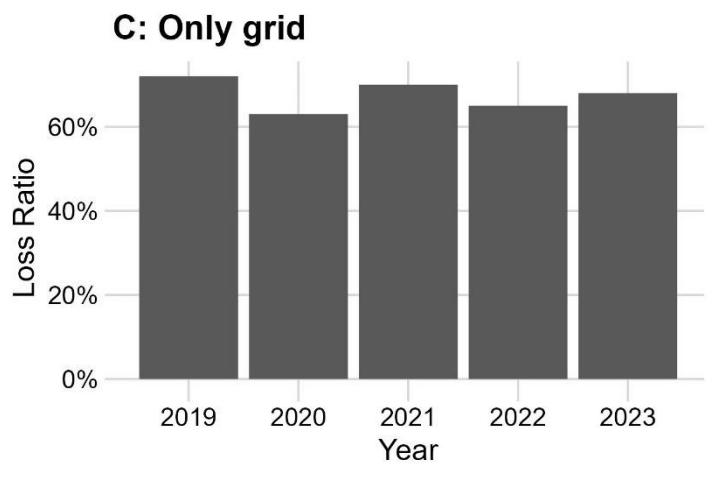
Labels - interactive

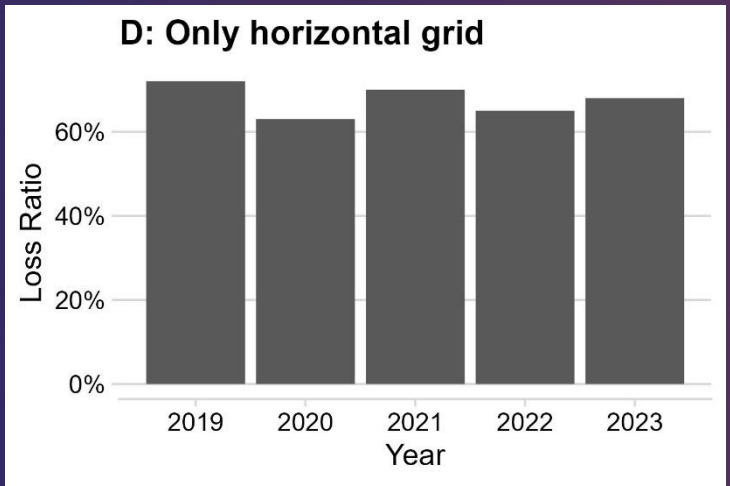


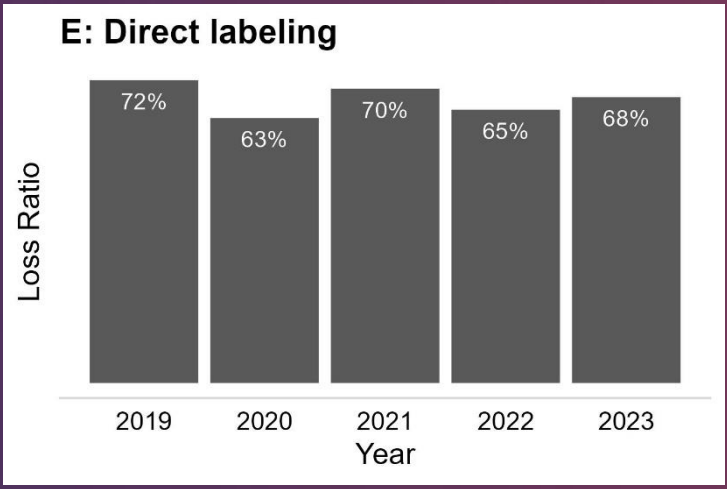
Poll: Which of the following charts appeals to you the most?











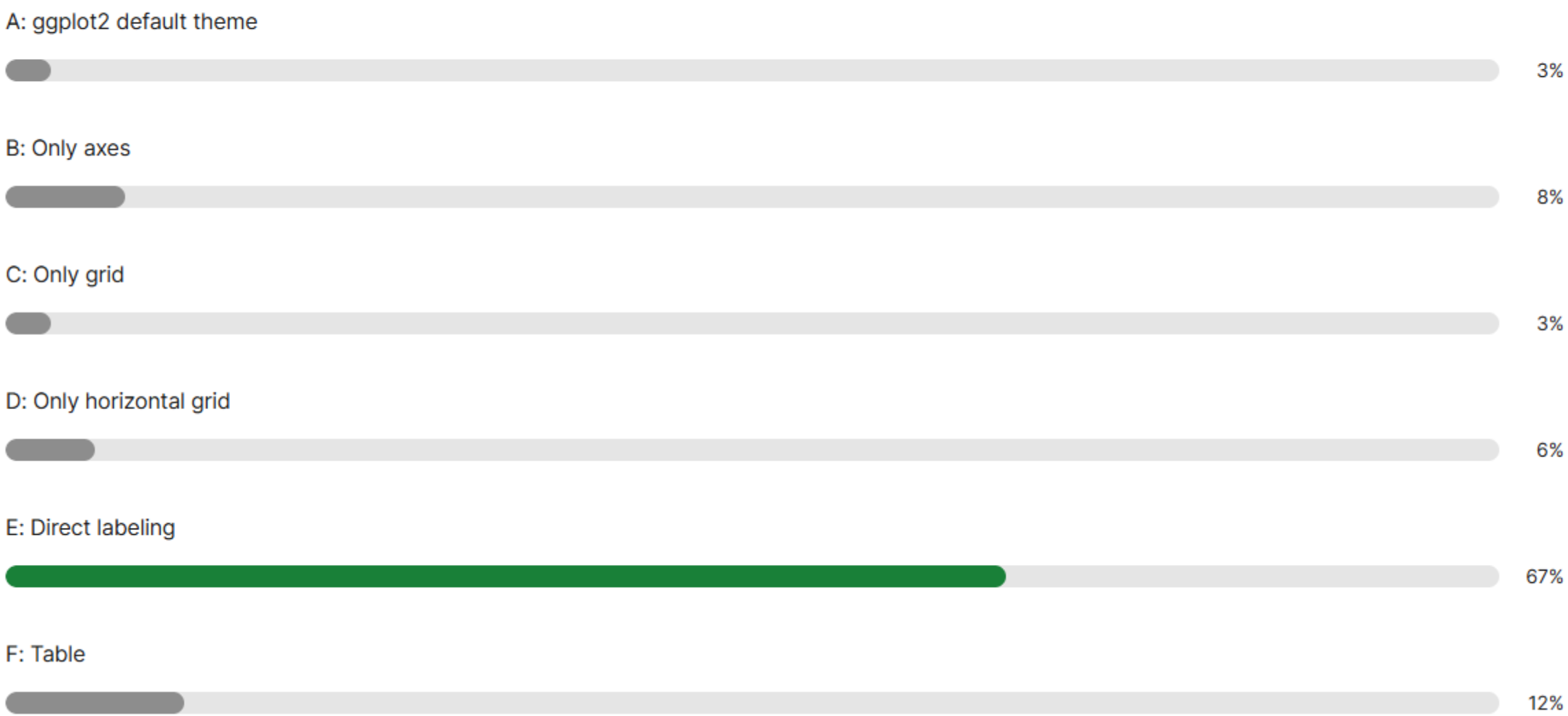
F: Table

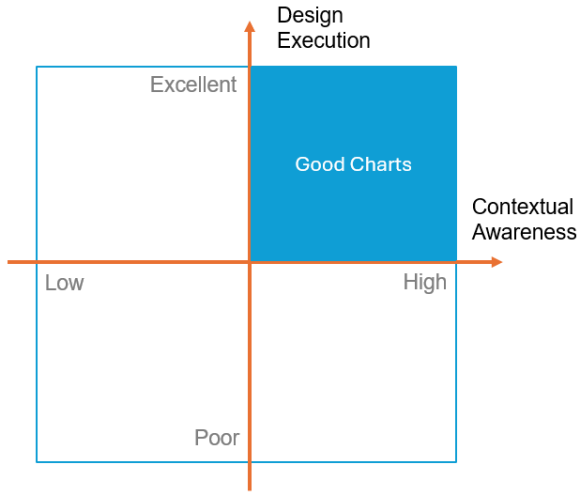
Year	Loss Ratio
2019	72%
2020	63%
2021	70%
2022	65%
2023	68%

Multiple choice
306 votes

Poll settings

Which of the following charts appeals to you the most?



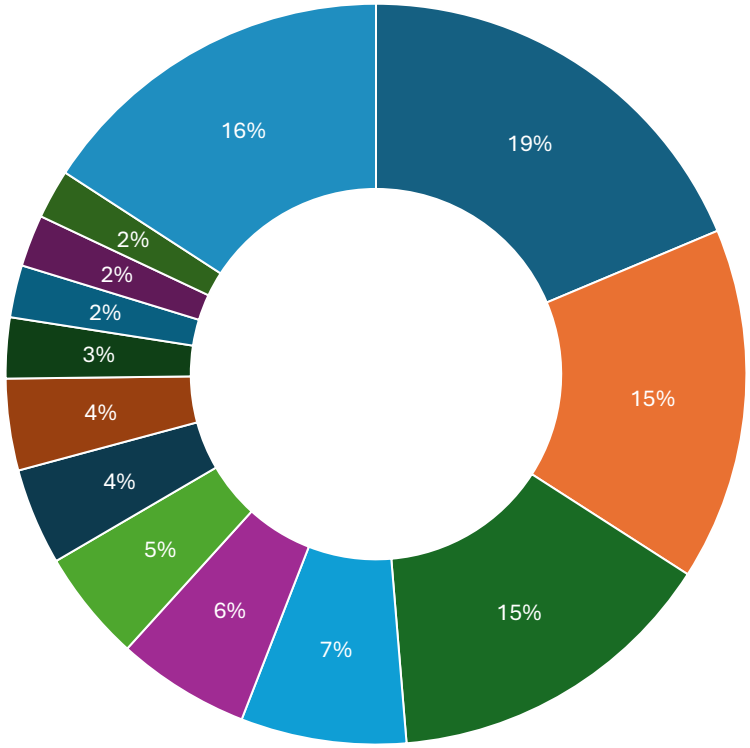


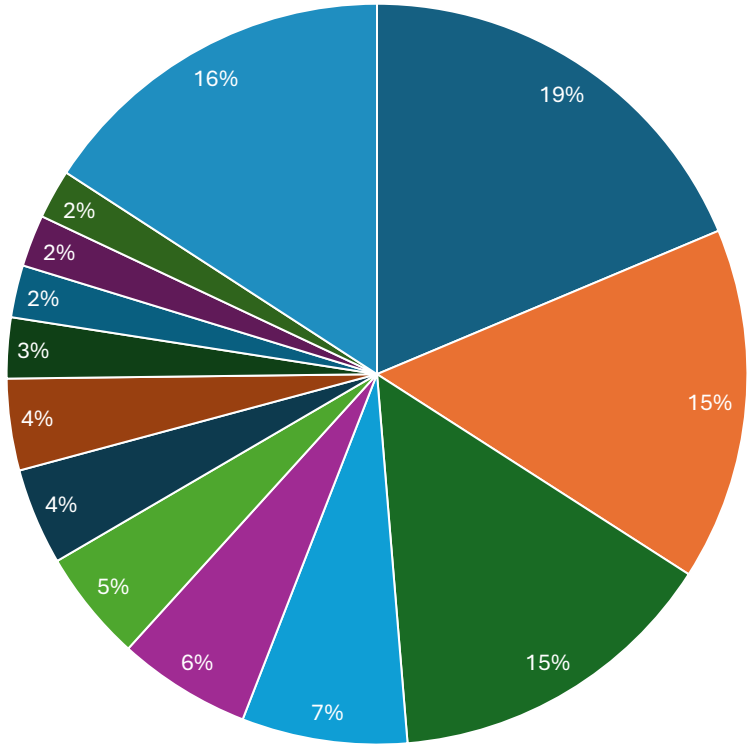
Contextual Awareness

Contextual Awareness

What?	Who?	When?	Where?	How?
• Amounts • Distributions • Proportions • Multivariate relationship • Geospatial • Uncertainty	• Actuaries • Non-actuaries (UW, brokers, management) • Regulators	• Data exploration phase • Modeling phase • Results communication phase	• Board meeting • Industry conference • Internal team standup call	• Oral presentation (verbally explained) • Written report (textually explained) • Infographics (self-explanatory)

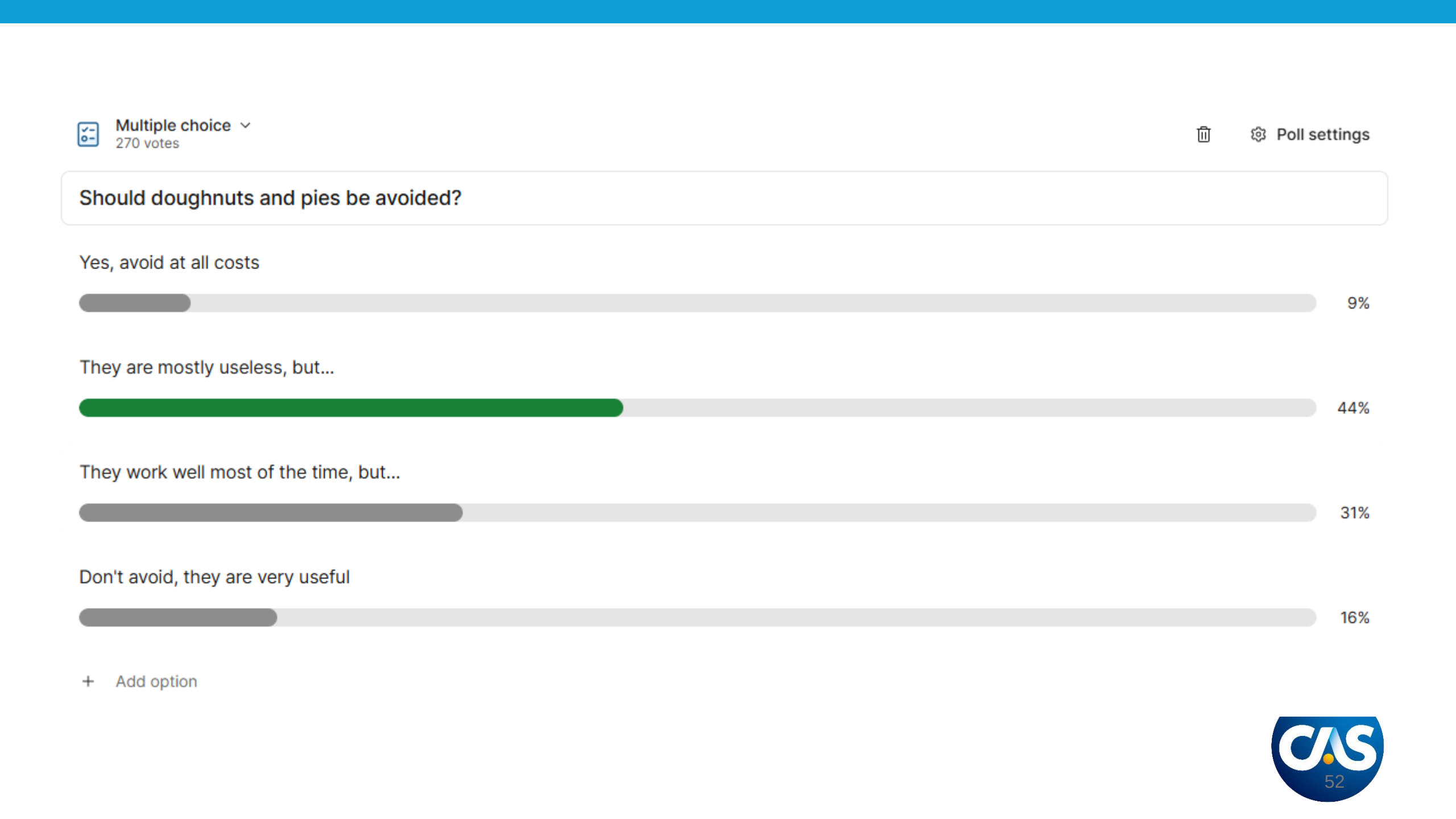
Is doughnut chart just pie chart with arbitrarily better data-ink ratio?





Poll: Should doughnuts and pies be avoided?

A. Yes, avoid at all costs
B. They are mostly useless, but…
C. They work well most of the time, but…
D. Don’t avoid, they are very useful


Multiple choice
270 votes
Poll settings
Should doughnuts and pies be avoided?
Yes, avoid at all costs
9%
They are mostly useless, but...
44%
They work well most of the time, but...
31%
Don't avoid, they are very useful
16%
Add option

You may strongly dislike pie charts…

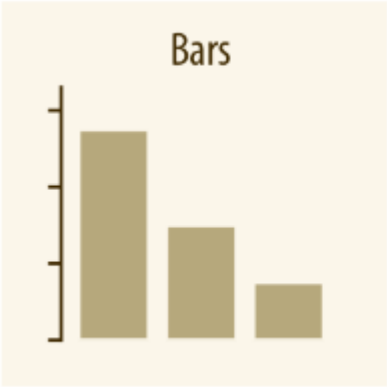
But if your principal loves pie charts, use pie charts despite your “visuarial” judgment.

“…taking into account of the intended users.” – ASOP No. 41

The “what”

Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

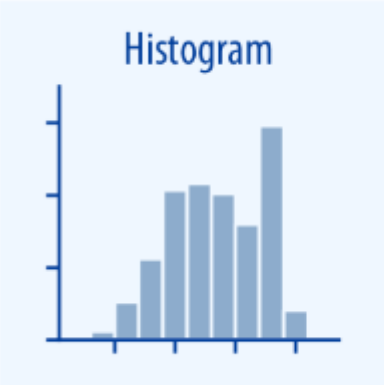
- Amounts

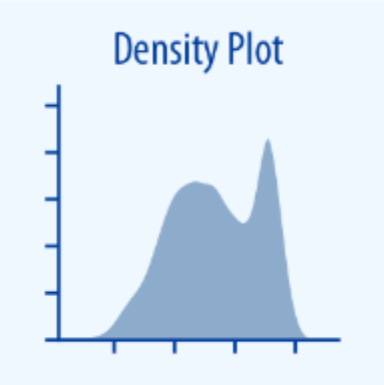


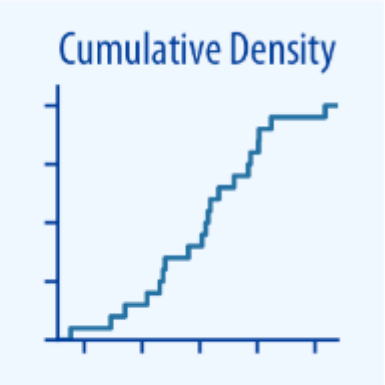
Fundamentals of Data Visualization (clauswilke.com)

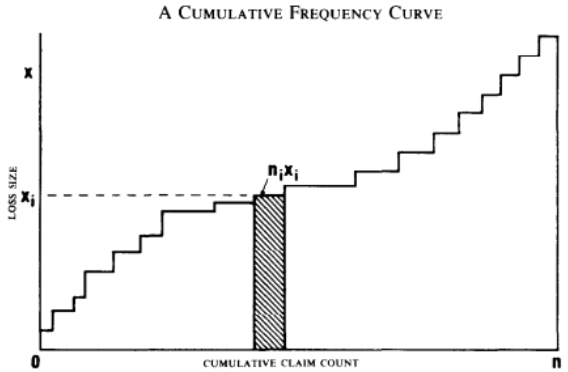
The “what”

- Single distribution









Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

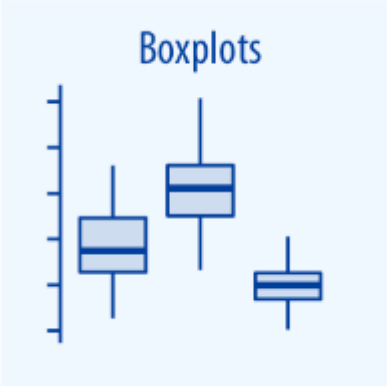
Fundamentals of Data Visualization (clauswilke.com)

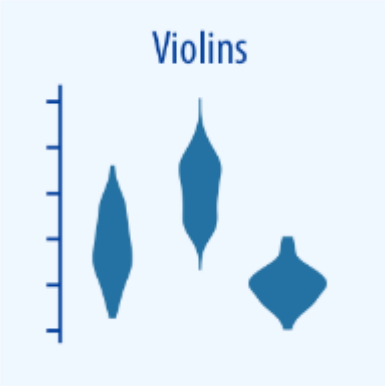
THE MATHEMATICS OF EXCESS OF LOSS COVERAGES AND RETROSPECTIVE RATING-A GRAPHICAL APPROACH (casact.org)

The “what”

Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

- Multiple distributions



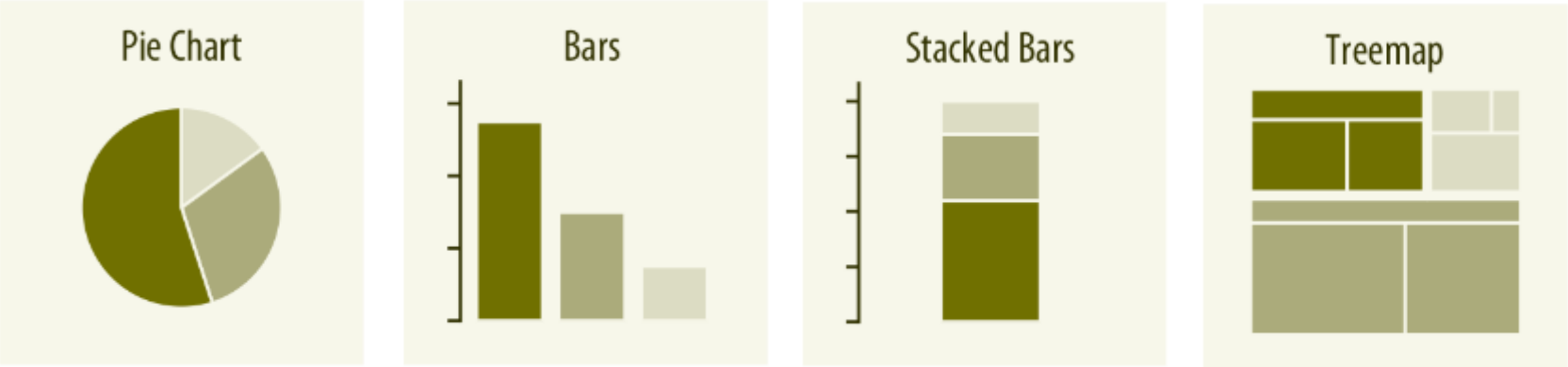


Fundamentals of Data Visualization (clauswilke.com)

The “what”

Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

- Proportions

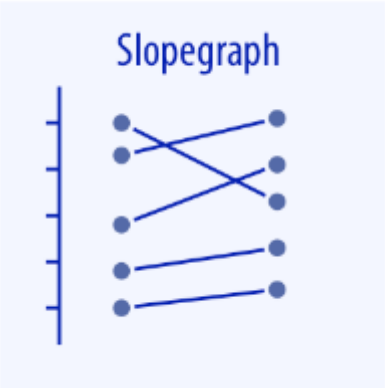


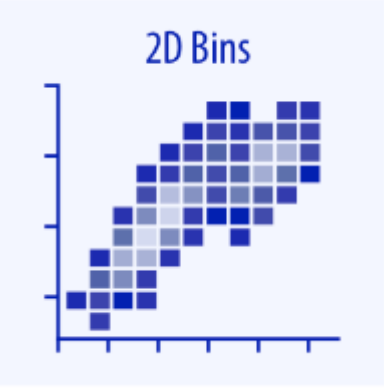
Fundamentals of Data Visualization (clauswilke.com)

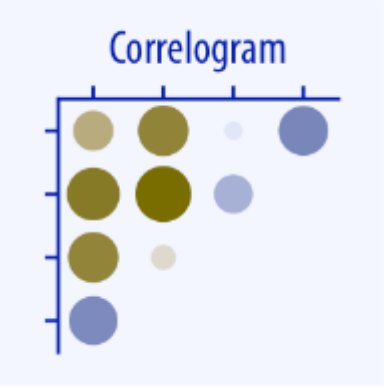
The “what”

Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

- Multivariate relationships





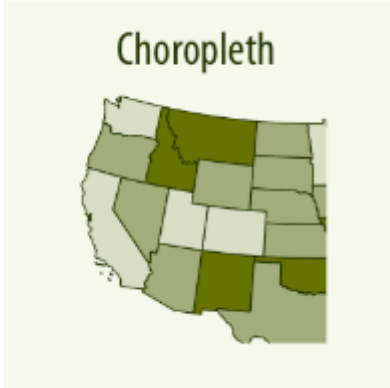


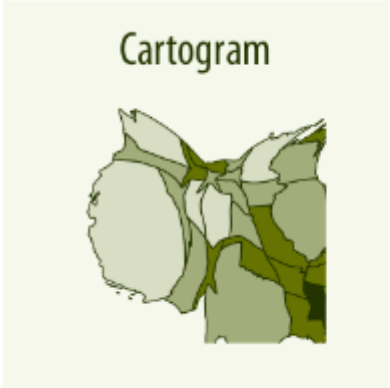
Fundamentals of Data Visualization (clauswilke.com)

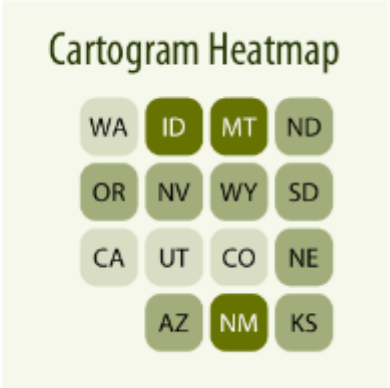
The “what”

Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.

- Geospatial





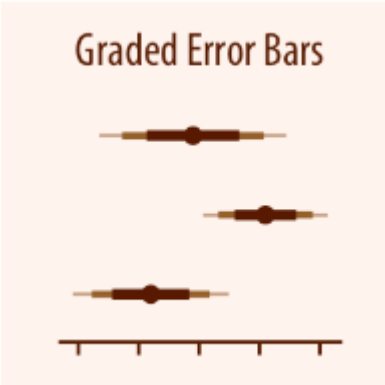


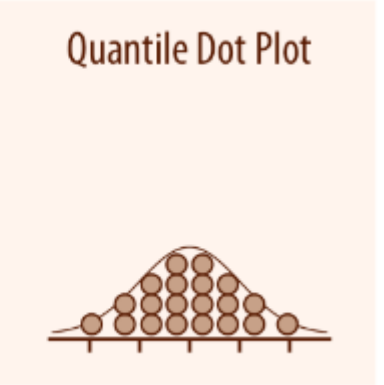
Fundamentals of Data Visualization (clauswilke.com)

The “what”

- Uncertainty

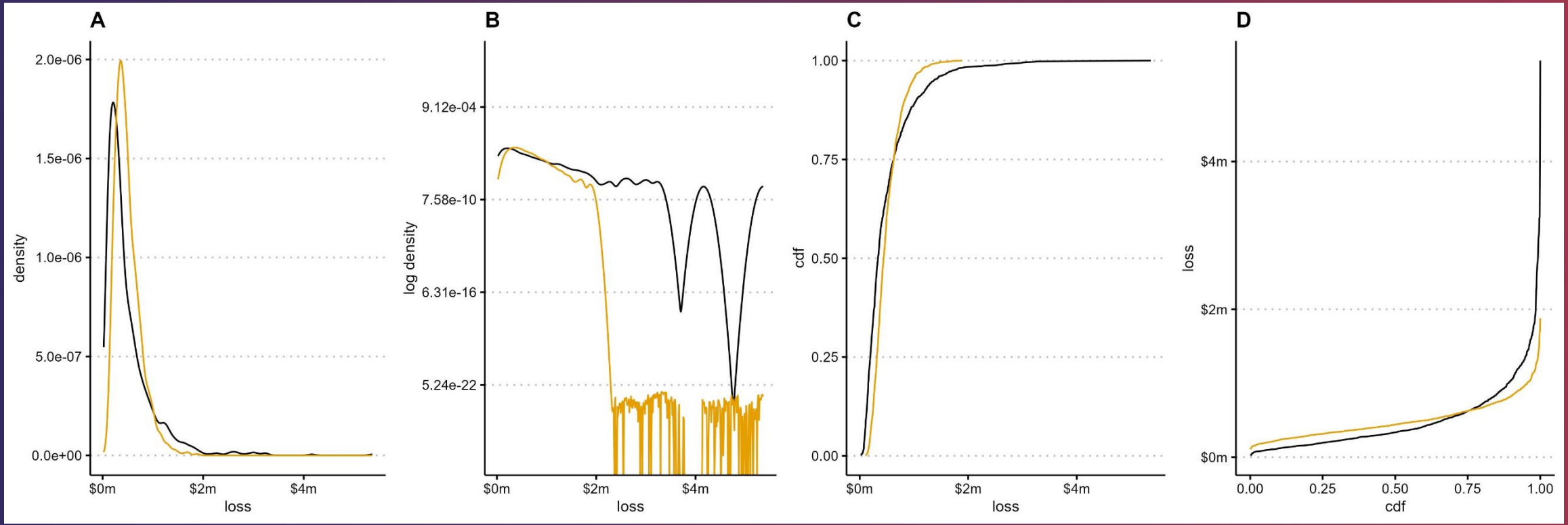
Practice thoughtfulness:
As we flip through these chart types, consider the “who, when, where, and how” to which each of these charts would be appropriate and effective.





Fundamentals of Data Visualization (clauswilke.com)

Poll: You want to communicate that a distribution is more heavy-tailed than the other. Which plot below would you choose for an actuarial audience?

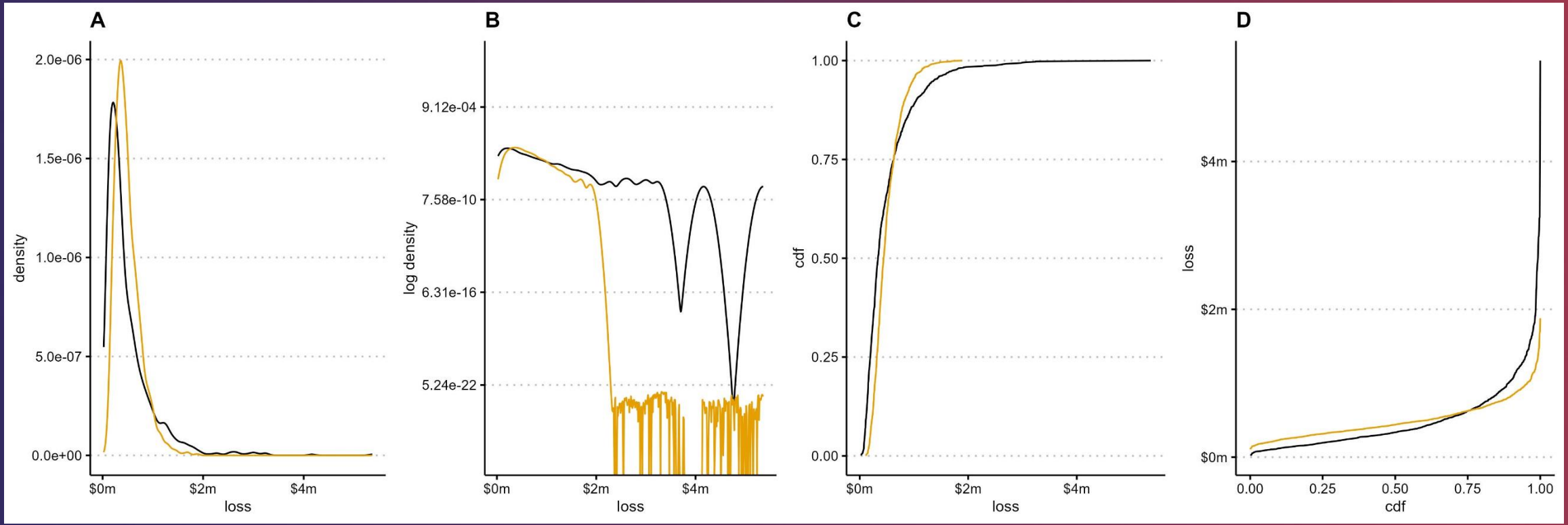


Multiple choice
246 votes

Poll settings

You want to communicate that a distribution is more heavy-tailed than the other. Which plot would you choose for an ACTUARIAL audience?

Poll: You want to communicate that a distribution is more heavy-tailed than the other. Which plot below would you choose for a non-actuarial audience?



Multiple choice
209 votes

Poll settings

You want to communicate that a distribution is more heavy-tailed than the other. Which plot would you choose for an NON-ACTUARIAL audience?

A: Density plot

32%

B: Log density plot

7%

C: CDF plot

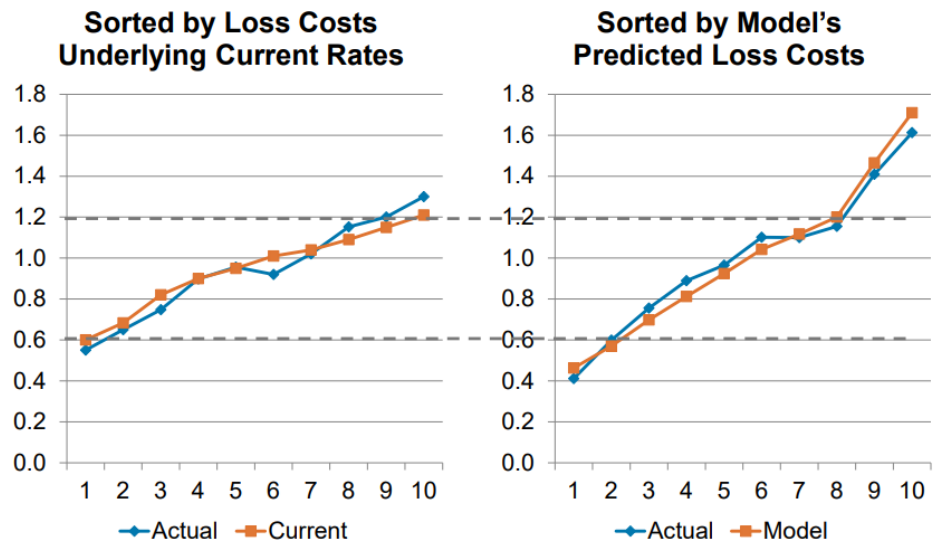
44%

D: Lee diagram

17%

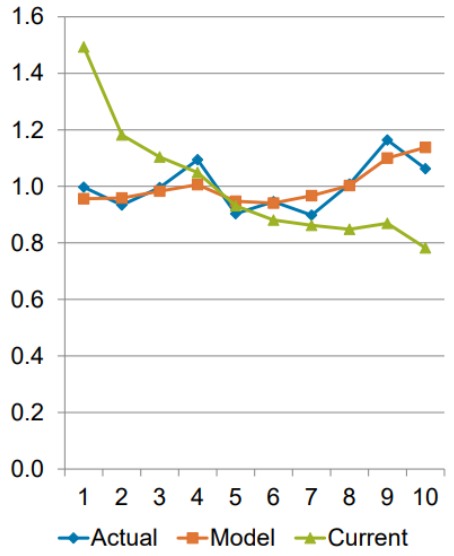
Poll: You want to communicate to management that your new predictive model is superior at differentiating high-cost policyholders from the low-cost ones. Which plot would you choose?

Model Lift – Simple Quantile Plots



Model Lift – Double Lift Charts

- Creating a double lift chart
  - Sort data by ratio of model prediction to current premium.
  - Subdivide sorted data into quantiles with equal exposure.
  - For each quantile calculate average actual loss cost, average model predicted loss cost and the average loss cost underlying the current manual premium .
  - Index the quantile averages to the overall averages.



rpm_2016_presentations_pm-lm-4.pdf (casact.org)

Multiple choice
252 votes

Poll settings

You want to communicate to MANAGEMENT that your new predictive model is superior at differentiating high-cost policyholders from the low-cost ones. Which plot would you choose?

Model Lift - Simple Quantile Plots

41%

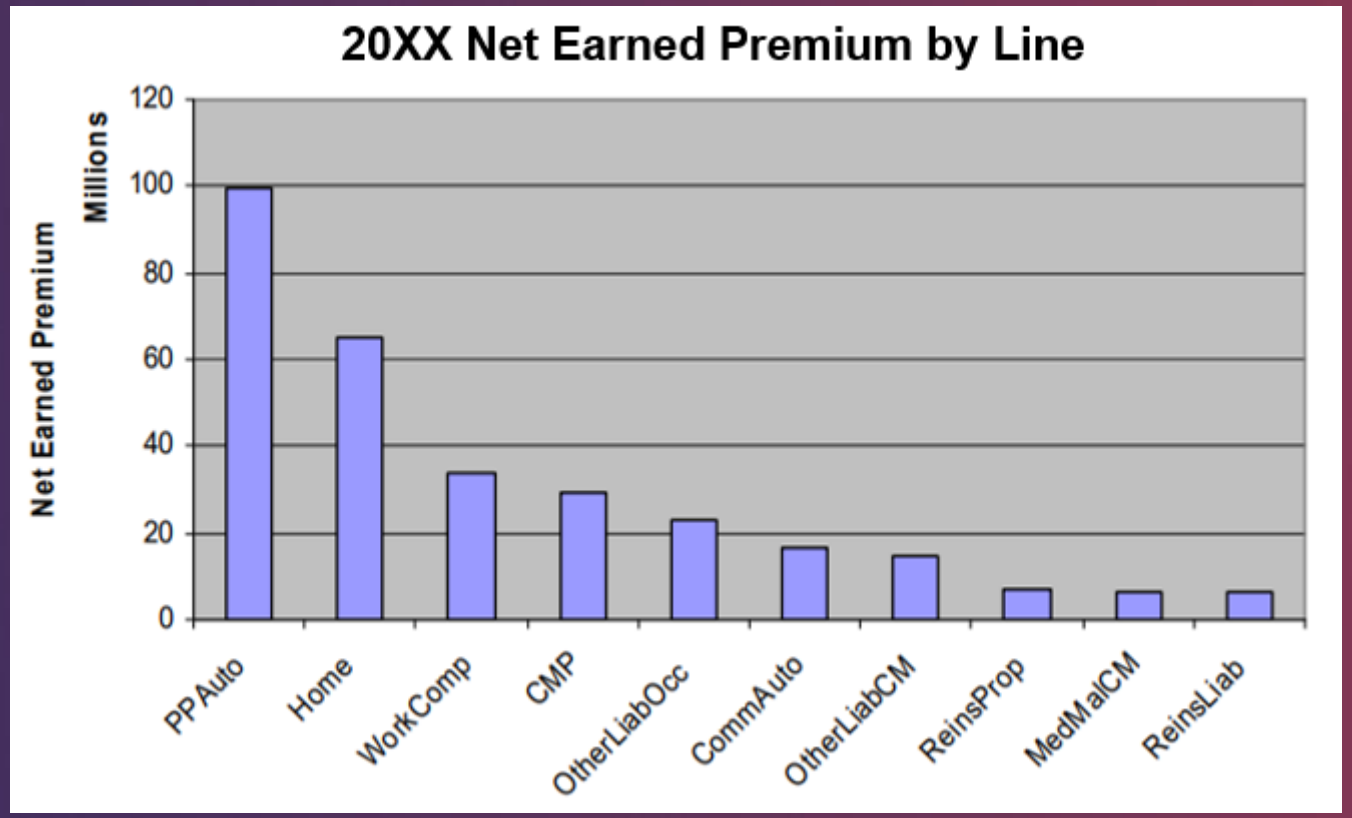
Model Lift - Double Lift Charts

59%

What I think of every time I see a double lift chart…

“Don’t design your visuals to be just clever or different. Design them so that they solve the problem in a materially improved, intuitive way.” – Jamie Mackay

So… how can we do better?



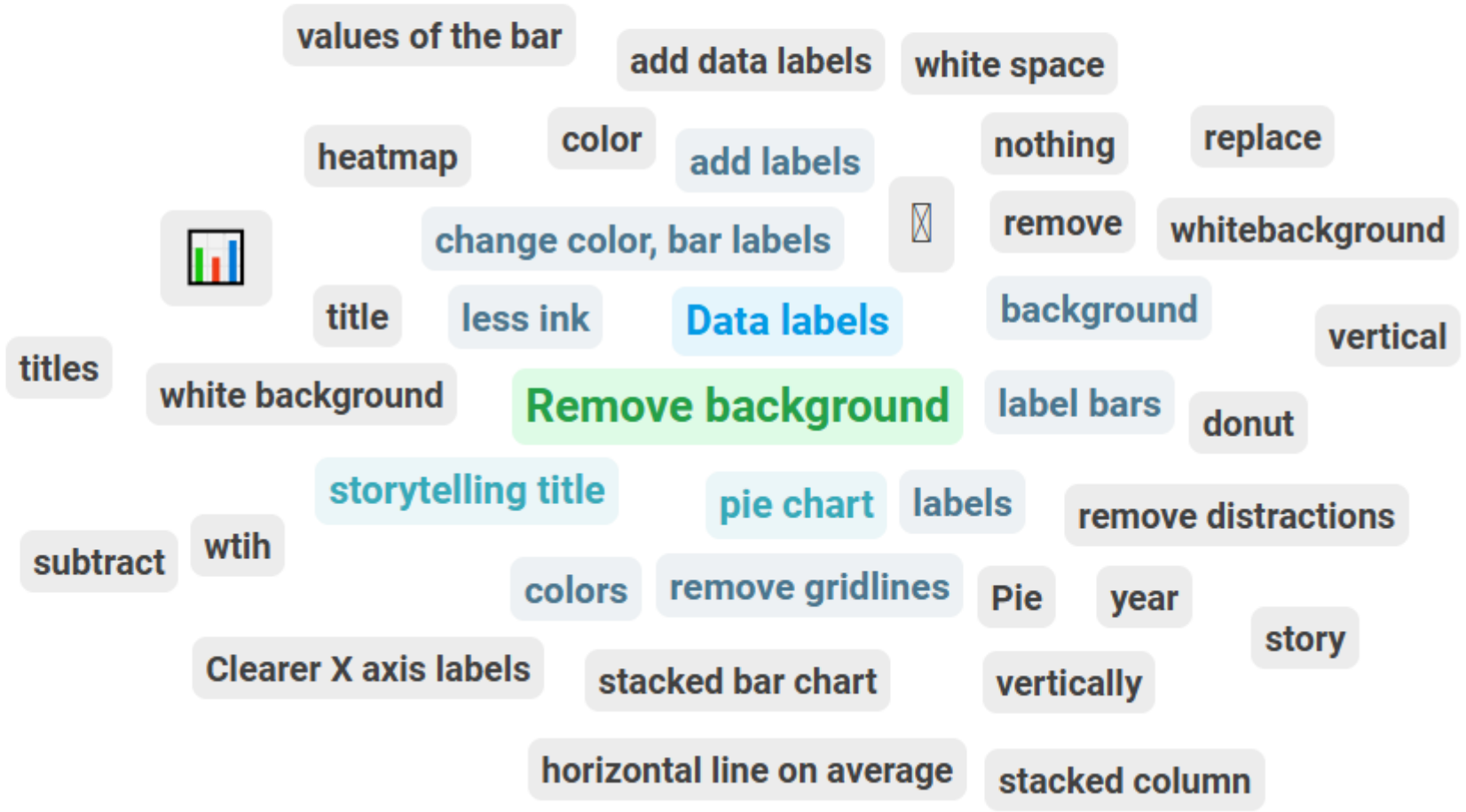
Word cloud
177 votes

Poll settings

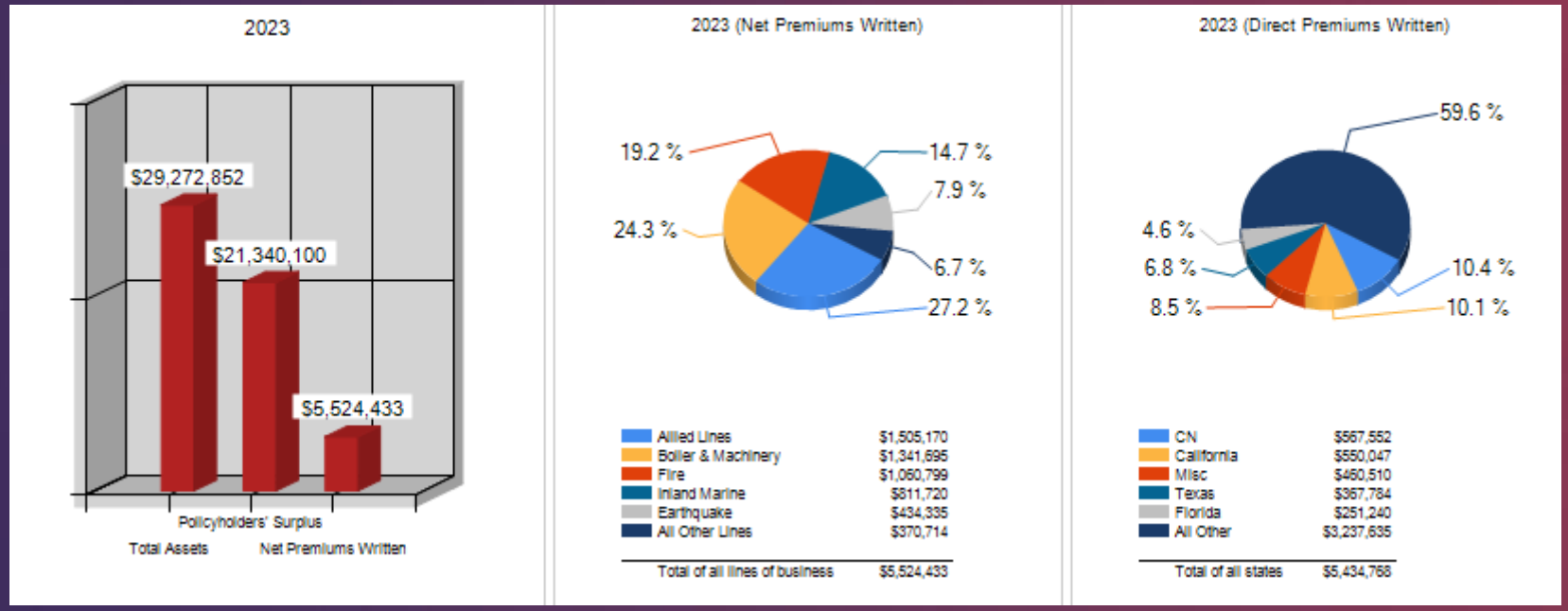
What can we do to improve this Net Earned Premium by Line chart?

Review answers

- Remove background 33
- Data labels 15
- pie chart 14
- Storytelling title 9
- change color, bar labels 7
- remove gridlines 5
- colors 4
- Add Labels 4



So… how can we do better?



Word cloud
178 votes

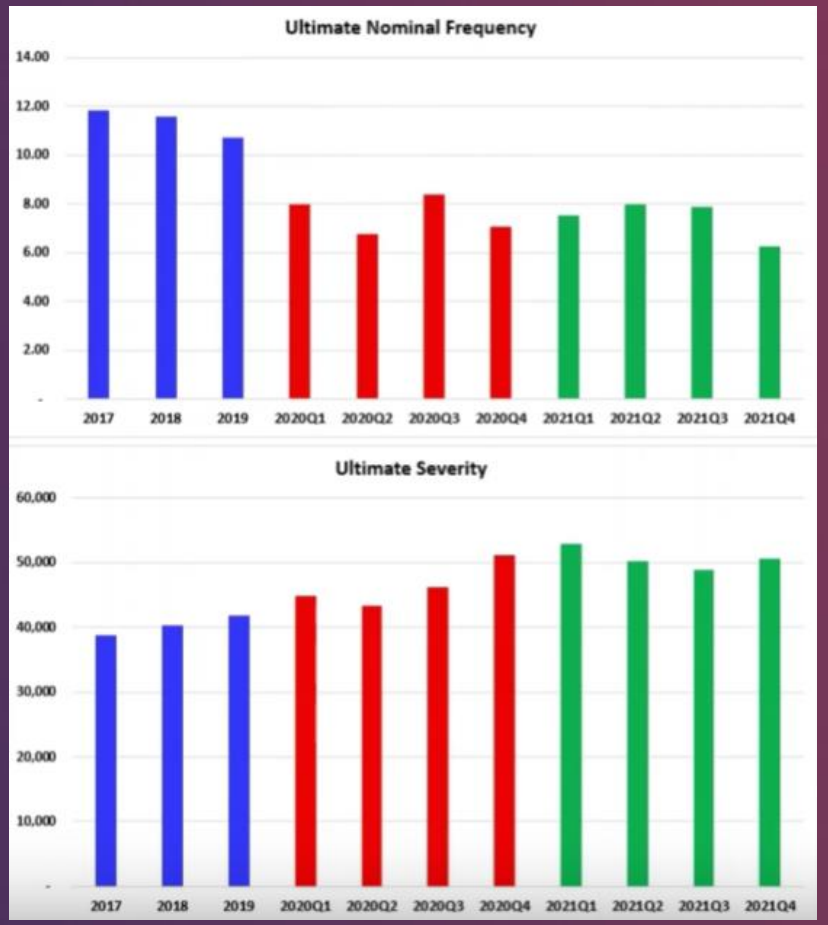
Poll settings

What can we do to improve this rating agency dashboard?

Review answers

- no 3D 24
- remove 3D 10
- Make 2D. 8
- no pie charts 7
- no 3D and no pie chart 6
- remove 3D effect 6
- storytelling title 5
- larger size 5

So… how can we do better?



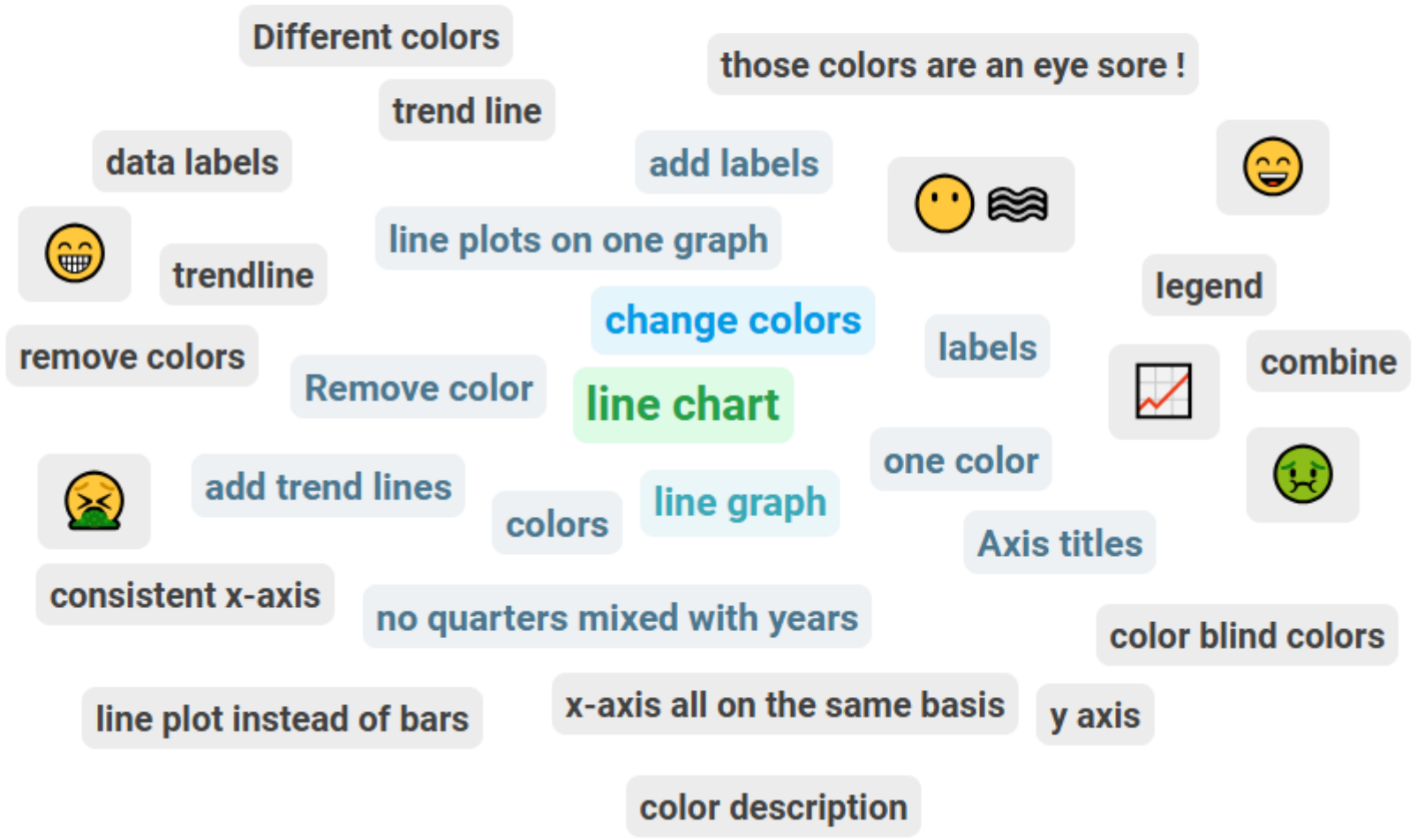
Word cloud
170 votes

Poll settings

What can we do to improve this frequency-severity "bi-plot"?

Review answers

- line chart 25
- change colors 13
- line graph 9
- Remove color 6
- line plots on one graph 4
- colors 4
- one color 4
- add trend lines 3
- labels 3



Conclusion – Be a ***Visuary***

1. We should value good data visualization
2. We can apply well researched design principles to improve the quality of charts
3. We need to be contextually aware and thoughtful when choosing and designing visualization

Poll: Out of everything that we discussed, what was the most impactful to you?

Word cloud
154 votes

Poll settings

Out of everything that we discussed, what was the most impactful to you?

Review answers

- contextual awareness 11
- Story-telling Titles 10
- Data:Ink ratio 9
- color blindness? never thought of it before 5
- examples 5
- simplicity of the visual 4
- less ink 4
- know your audience 4

Thank you!

- enbo.jiang@lockton.com
- https://www.linkedin.com/in/enbojiang/
- https://github.com/enbojiang/cas-talks

This Photo by Unknown Author is licensed under CC BY-NC