



2024 Spring Meeting

# ATL | **Adapt Transform Lead**

May 5-8 2024 • Atlanta, Georgia

# 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.

# Be a Visuary

Data Visualization for Actuaries

2024.05.07

Enbo Jiang, FCAS, CERA



# 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 visualizations
- E. What is data?!

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

- A. Spreadsheet
- B. The likes of Power BI / Tableau
- C. R ggplot & friends
- D. Python matplotlib & friends
- E. Paint (!)

# Introduction





# Not a new topic...

Death by Chartjunk? Graphical Excellence In Insurance

CAS 2012 Annual Meeting

Data Visualization

Visual Thinking

- Data Visualization for Actuaries and Data Scientists

The importance of principles

Keith Quigley, Actuary

Picture This:  
Using Data Visualization

A graph is worth a thousand words

The effective use of visualization

CLRS 2022 St. Louis

Jamie Mackay

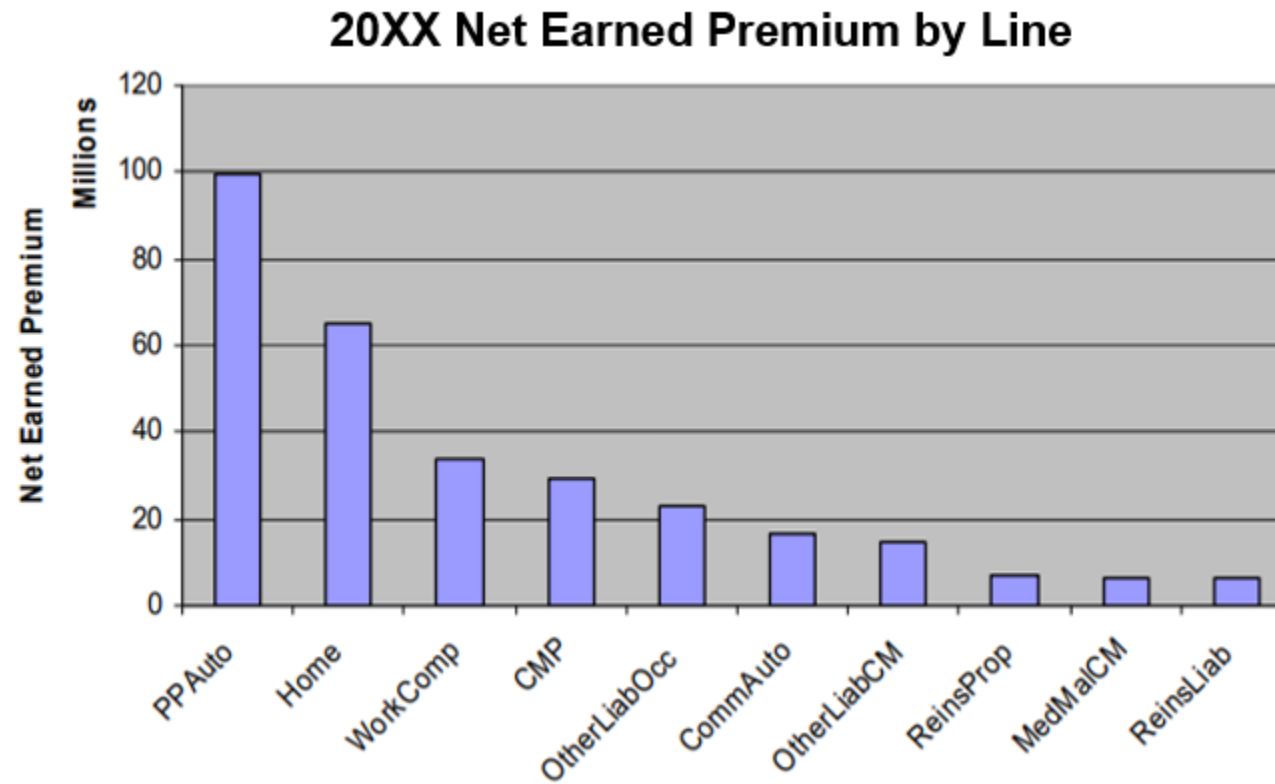
## Introduction to Data Visualization

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

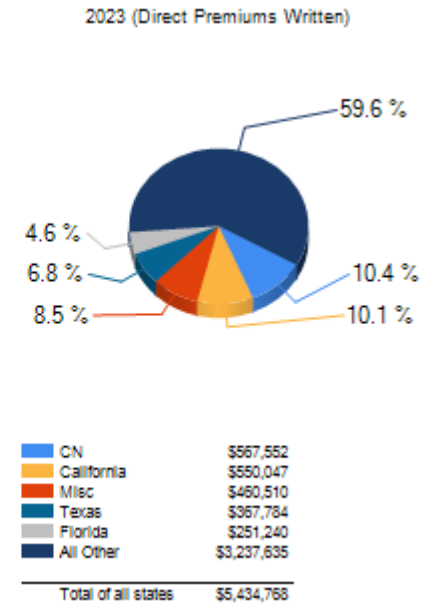
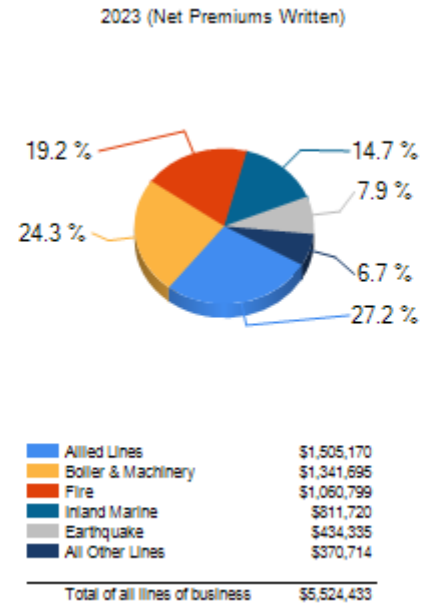
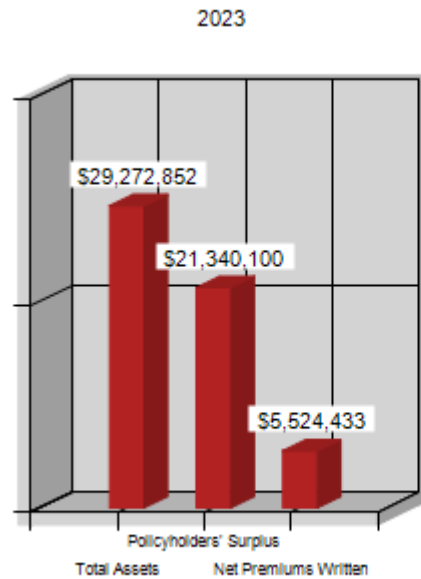
# But still seeing this?



Annual Statement Data



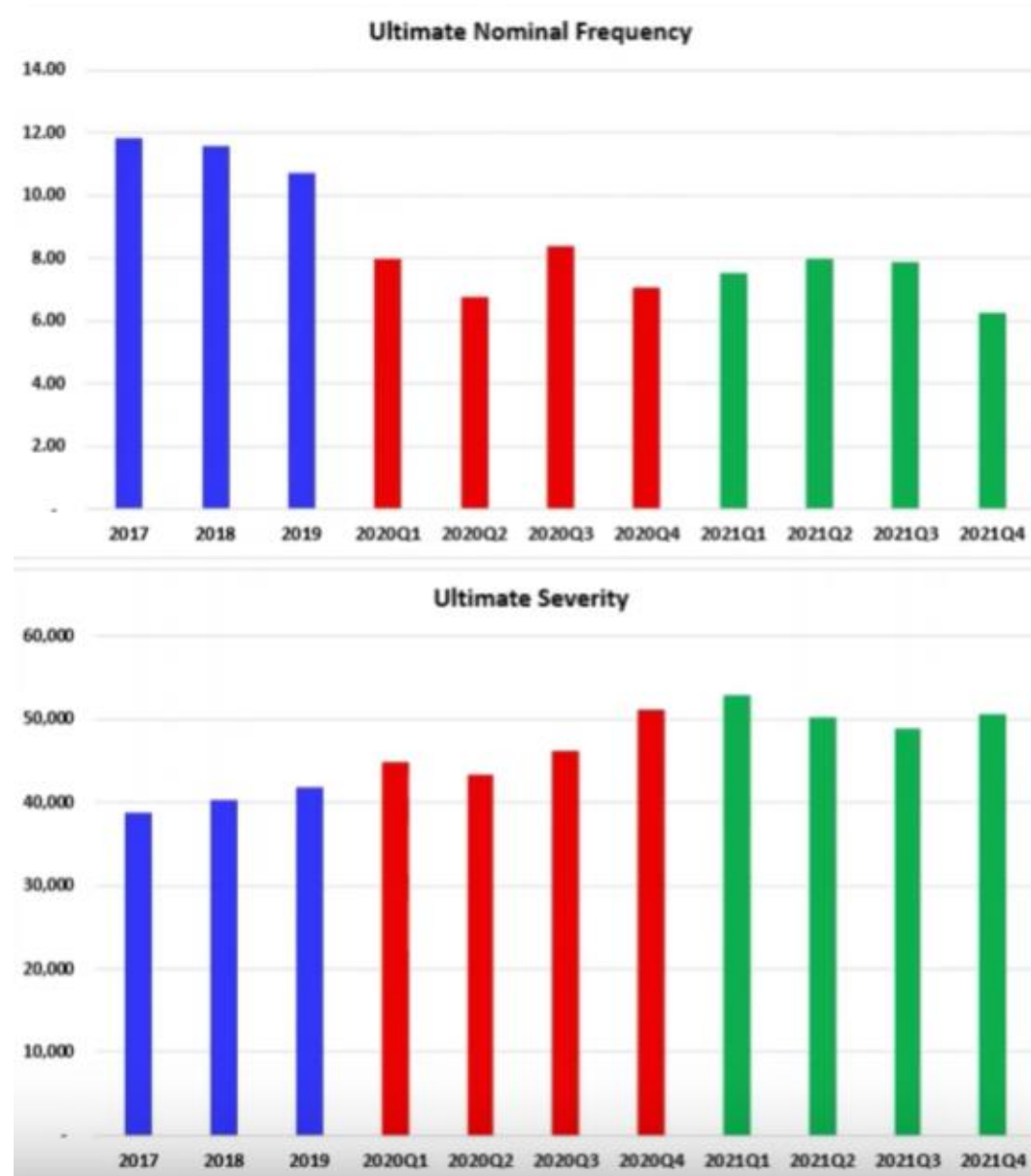
# Or this?



An unnamed rating agency



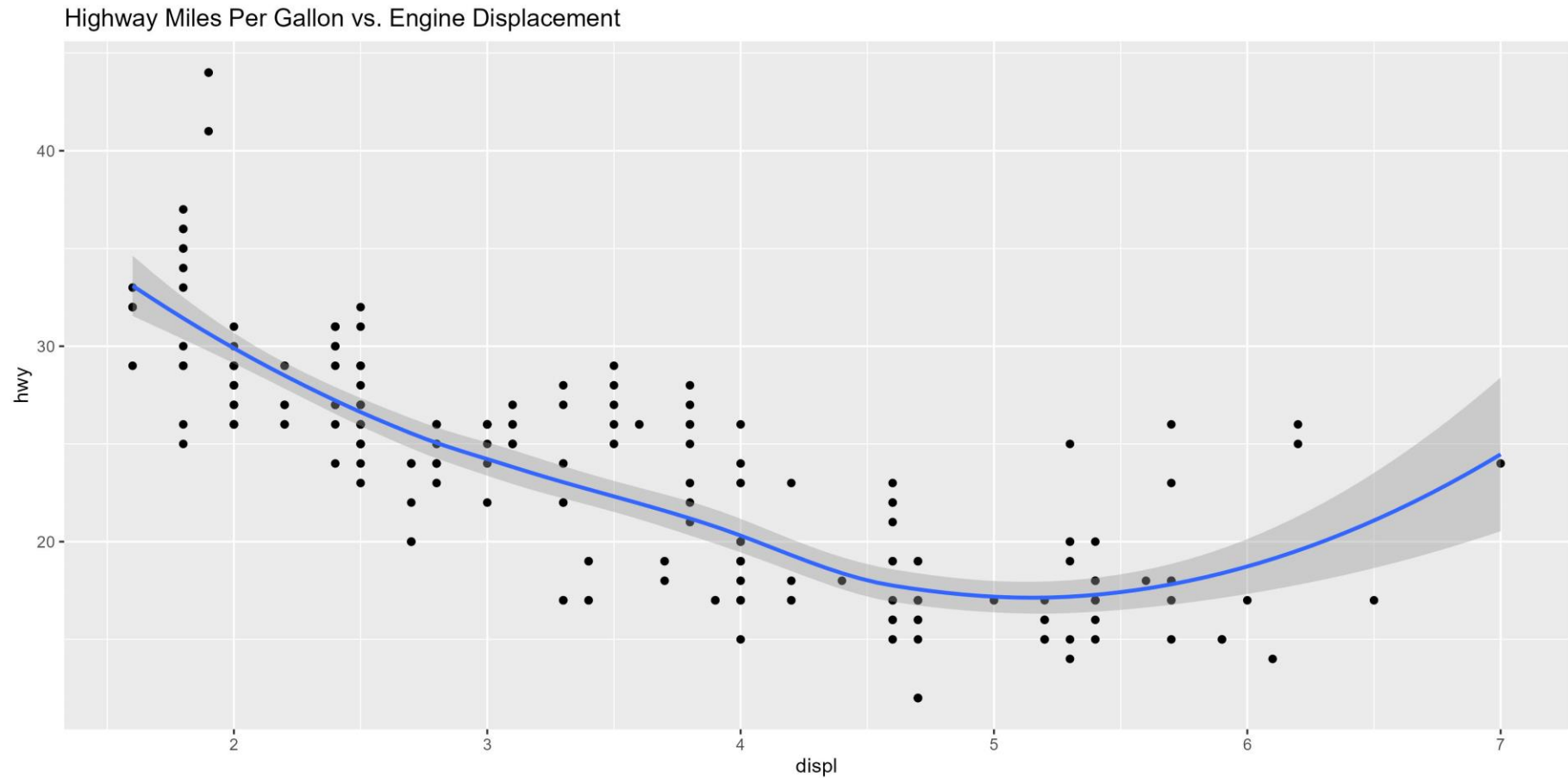
# Or this?



An unnamed industry  
benchmark provider

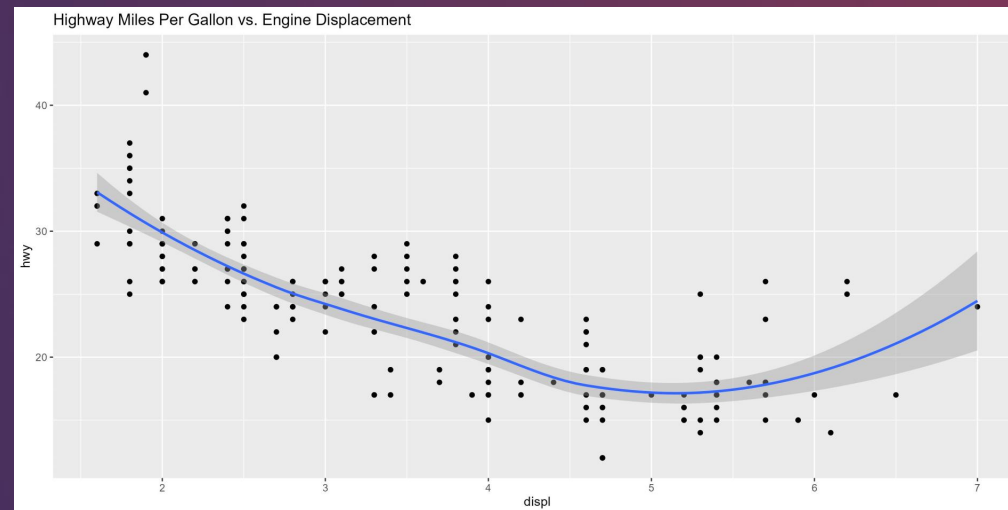
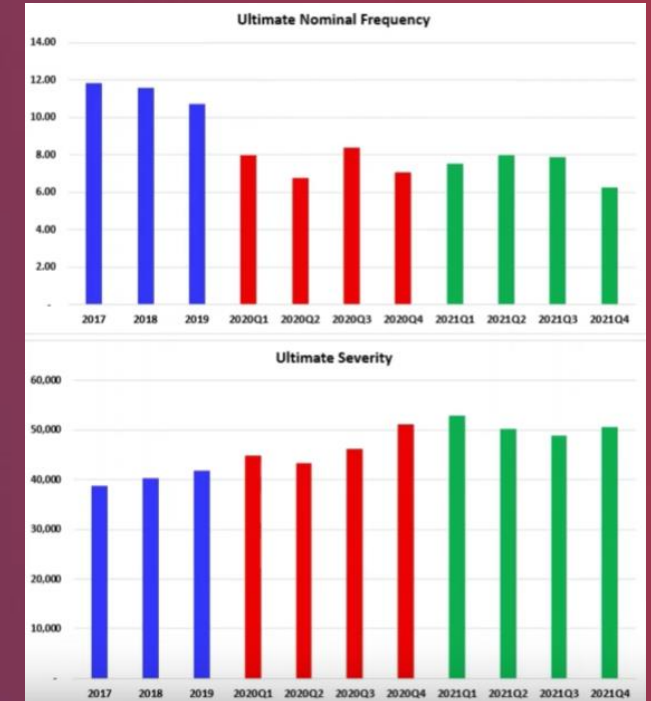
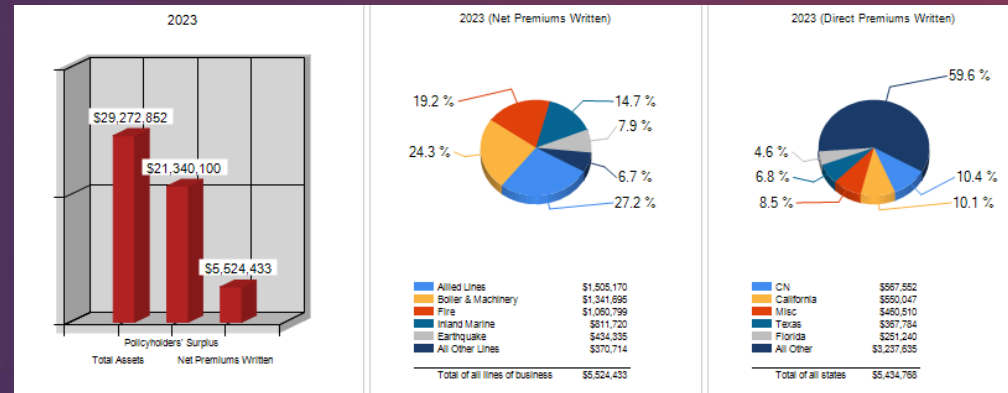
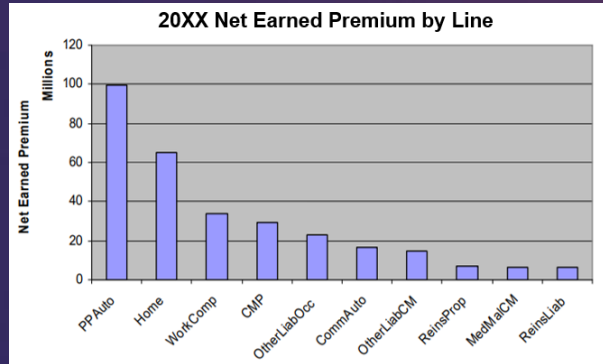


# What about this?



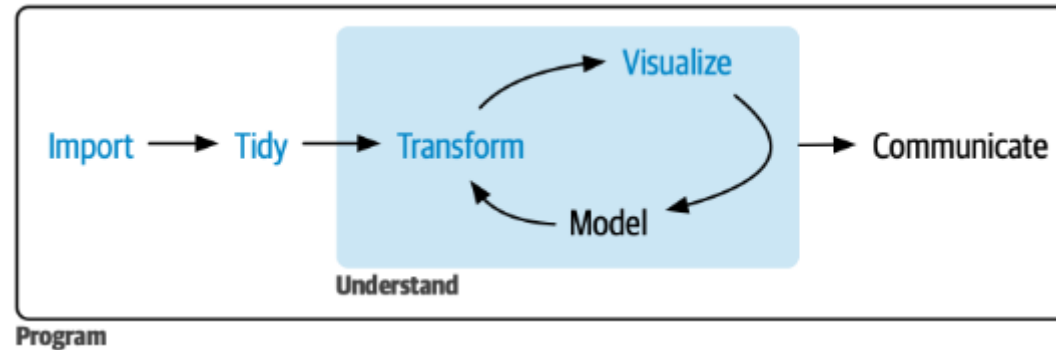
<https://ggplot2.tidyverse.org/reference/mpg.html>

# How can we do better?



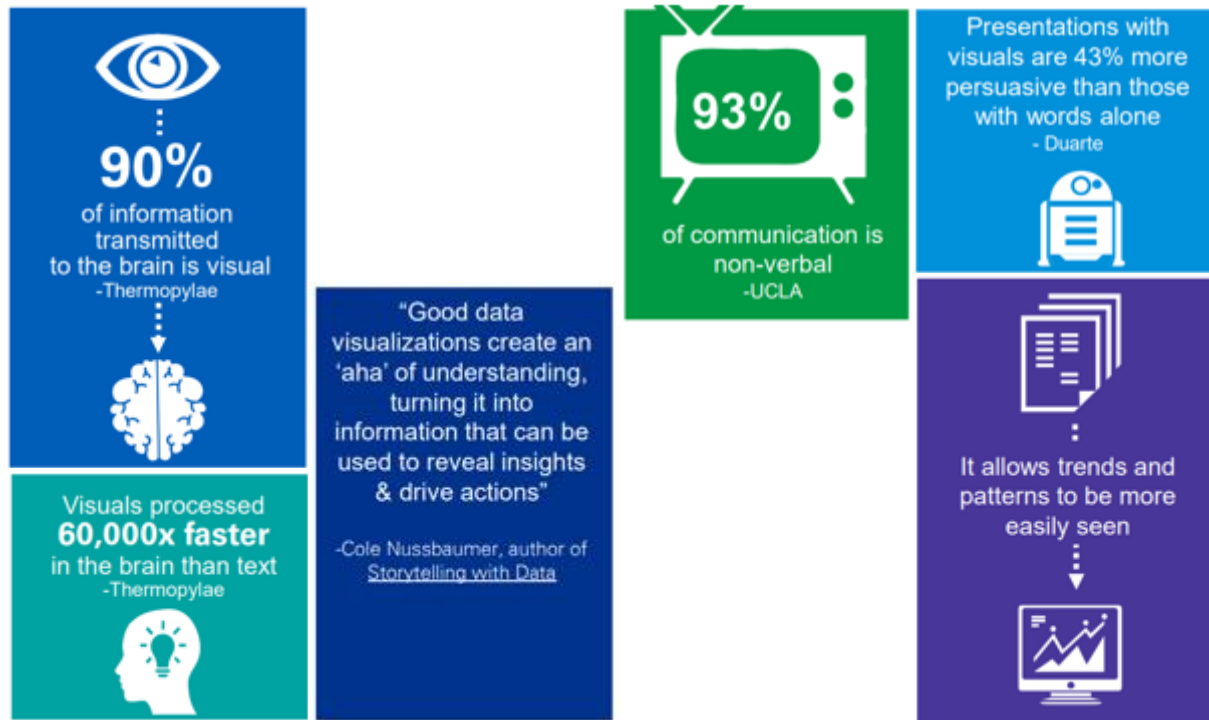
# Why should we care?

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



# Why should we care?

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





# 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

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

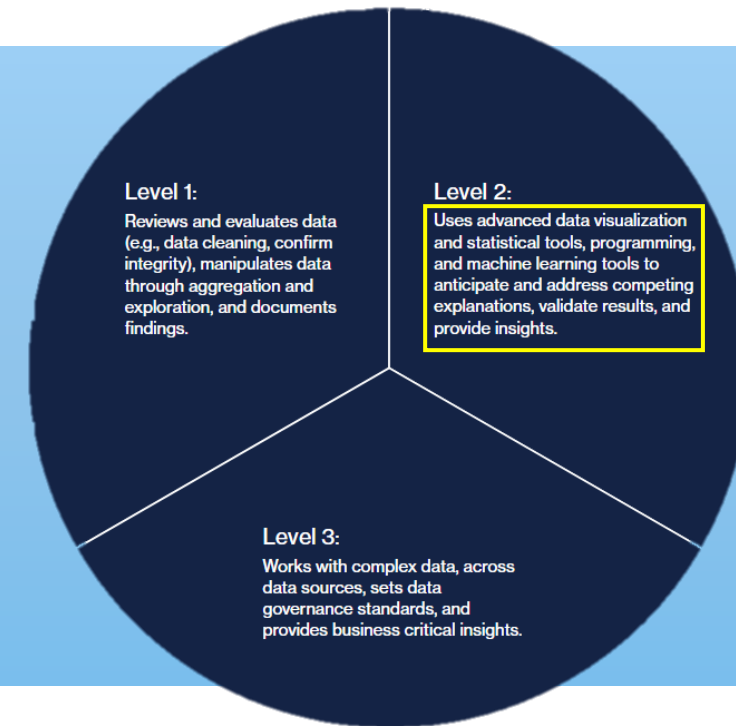
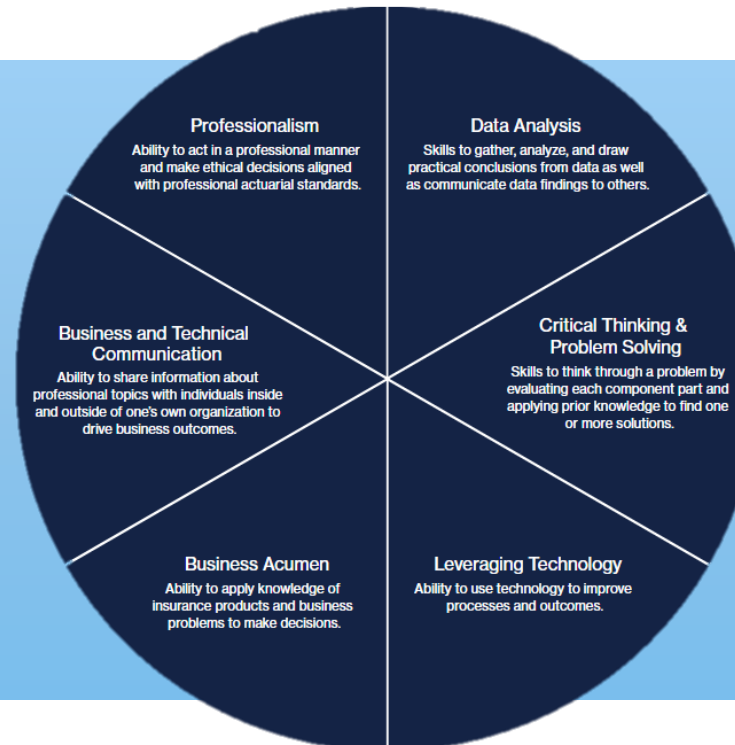
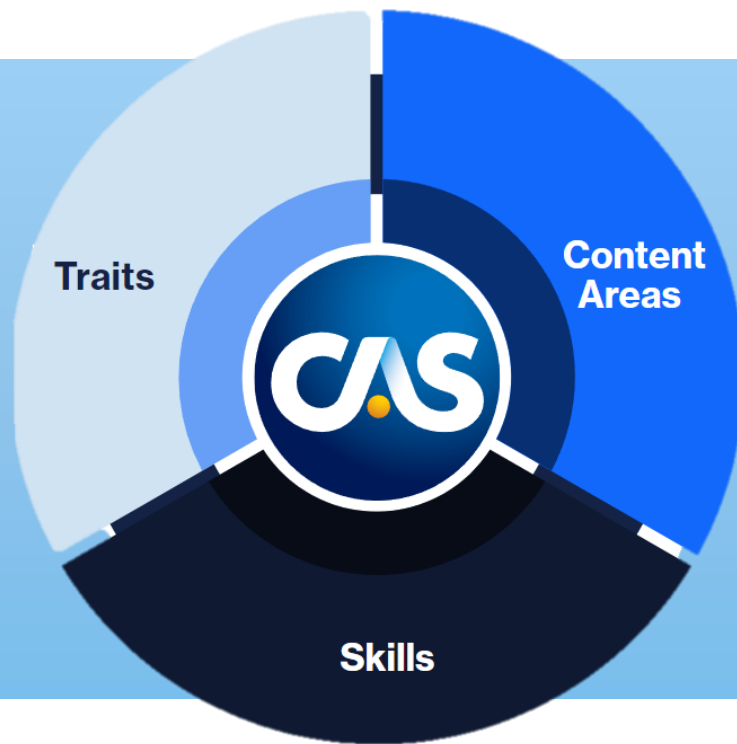
- Data visualization wasn't part of the exam syllabus at all until very recently through the new online course

## 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 Capability Model



# 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

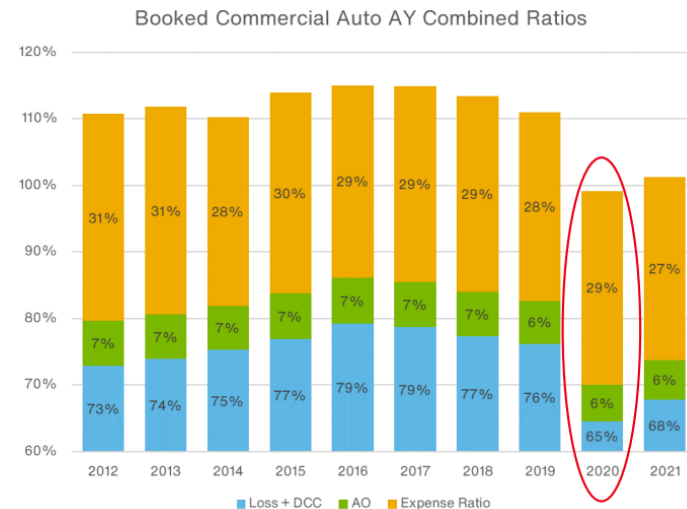
# Tufte's Graphical Excellence

- 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
- Graphical excellence requires telling the **truth**
- ...
- “...above all else, show the data”

# Good viz is not misleading

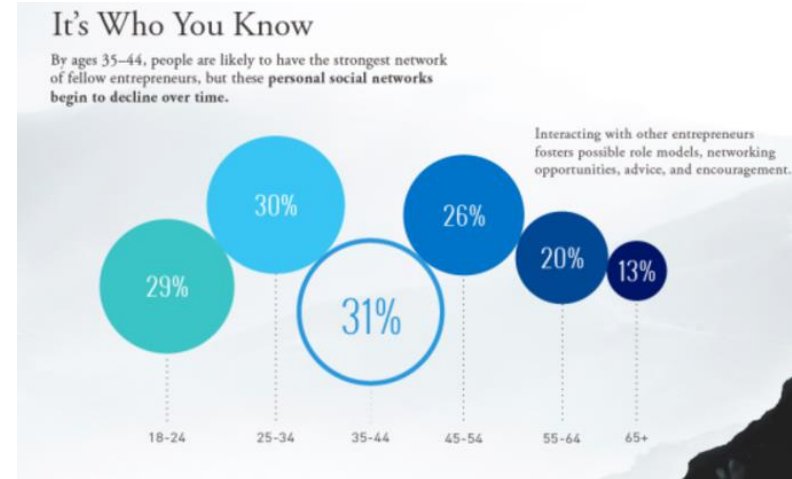
- “Graphical excellence requires telling the truth about the data.”
  - Visual Display of Quantitative Information pg. 51

## Obvious



PowerPoint Presentation (casact.org)

## Subtle

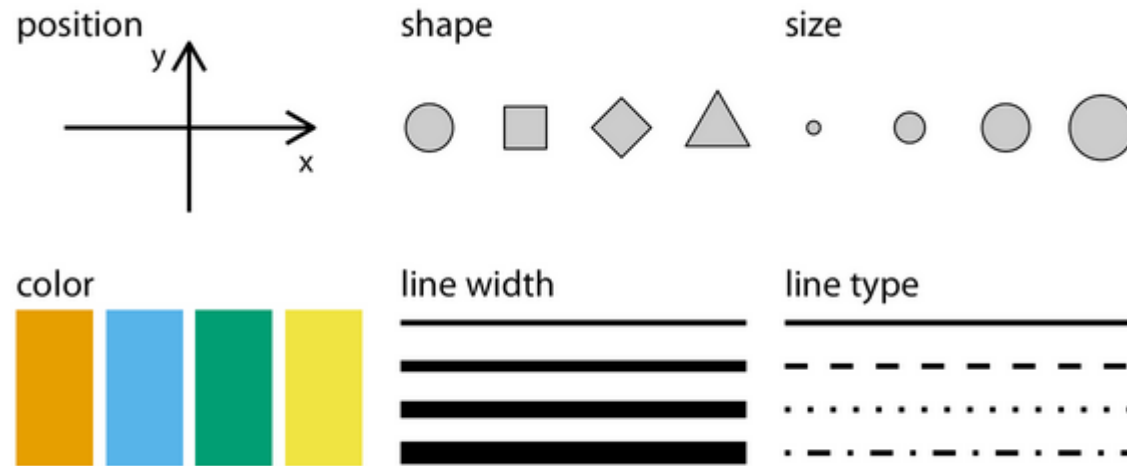


Tools - Proportional Ink (callingbullshit.org)



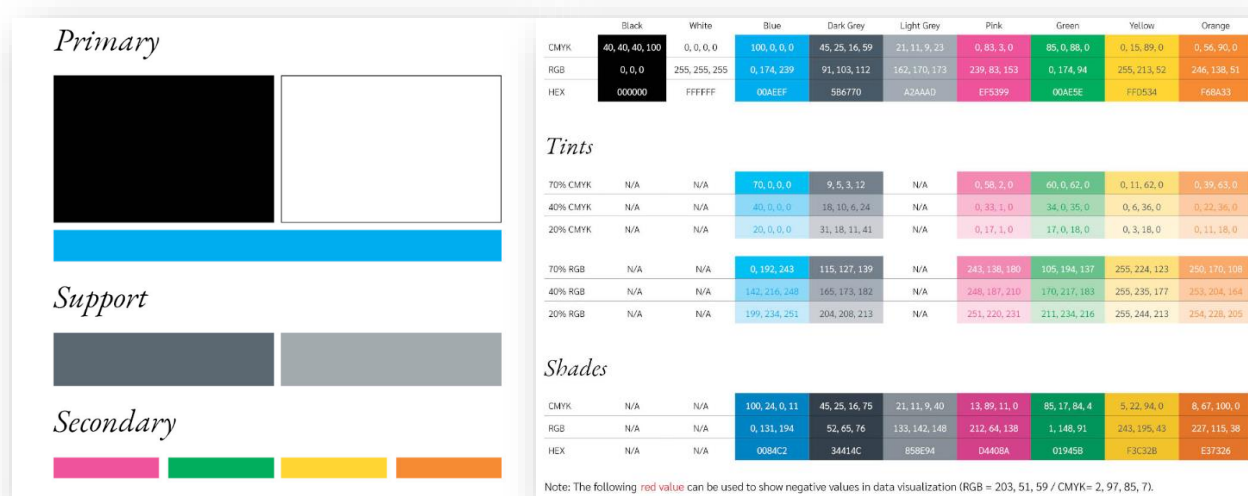
# Good viz employs intentional aesthetics

- “Aesthetics”
  - *aes()* in *ggplot*
  - Combine color with others 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!



*Primary*

*Support*

*Secondary*

	Black	White	Blue	Dark Grey	Light Grey	Pink	Green	Yellow	Orange
CMYK	40, 40, 40, 100	0, 0, 0, 0	100, 0, 0, 0	45, 25, 16, 59	21, 11, 9, 23	0, 83, 3, 0	85, 0, 88, 0	0, 15, 89, 0	0, 56, 90, 0
RGB	0, 0, 0	255, 255, 255	0, 174, 239	91, 103, 112	162, 170, 173	239, 83, 153	0, 174, 94	255, 213, 52	246, 138, 51
HEX	000000	FFFFFF	00AEF7	5B677D	A2A9AD	EF5399	00AE5E	FFD534	F6BA33

*Tints*

	70% CMYK	40% CMYK	20% CMYK	70% RGB	40% RGB	20% RGB
Blue	N/A	N/A	N/A	N/A	N/A	N/A
Dark Grey	N/A	N/A	N/A	N/A	N/A	N/A
Light Grey	N/A	N/A	N/A	N/A	N/A	N/A
Pink	N/A	N/A	N/A	N/A	N/A	N/A
Green	N/A	N/A	N/A	N/A	N/A	N/A
Yellow	N/A	N/A	N/A	N/A	N/A	N/A
Orange	N/A	N/A	N/A	N/A	N/A	N/A

*Shades*

	CMYK	RGB	HEX
Blue	N/A	N/A	N/A
Dark Grey	N/A	N/A	N/A
Light Grey	N/A	N/A	N/A
Pink	N/A	N/A	N/A
Green	N/A	N/A	N/A
Yellow	N/A	N/A	N/A
Orange	N/A	N/A	N/A

Note: The following red value can be used to show negative values in data visualization (RGB = 203, 51, 59 / CMYK = 2, 97, 85, 7).

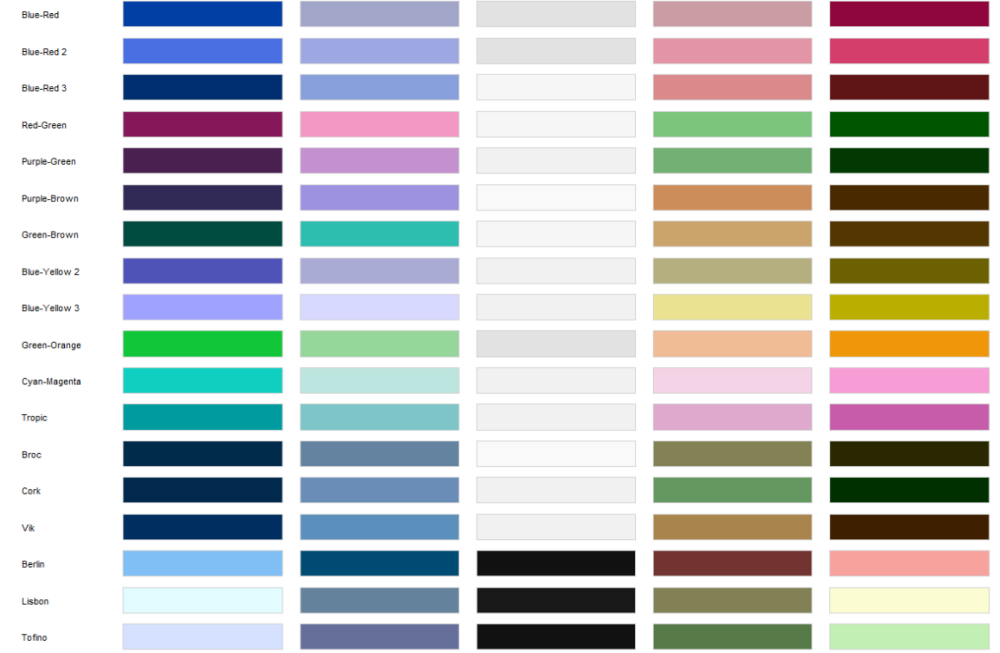
- If not, understand the use cases:
  - Categories (qualitative), numeric values (sequential/diverging), highlight

# Color Considerations

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



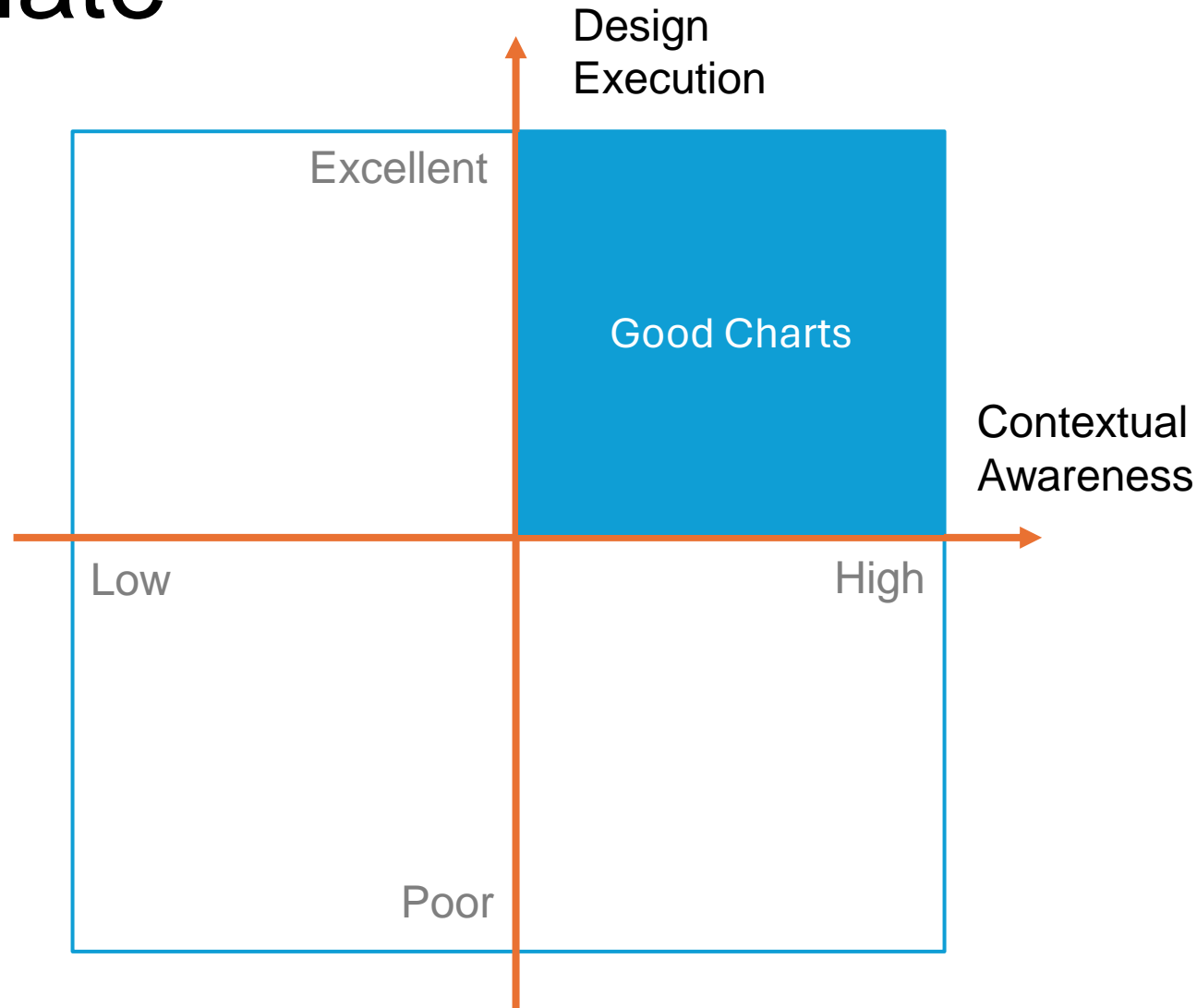
## Diverging

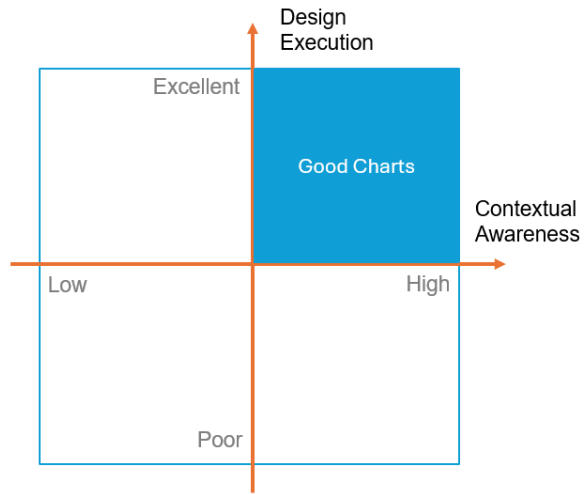


## Qualitative



# Good viz is well designed and context-appropriate

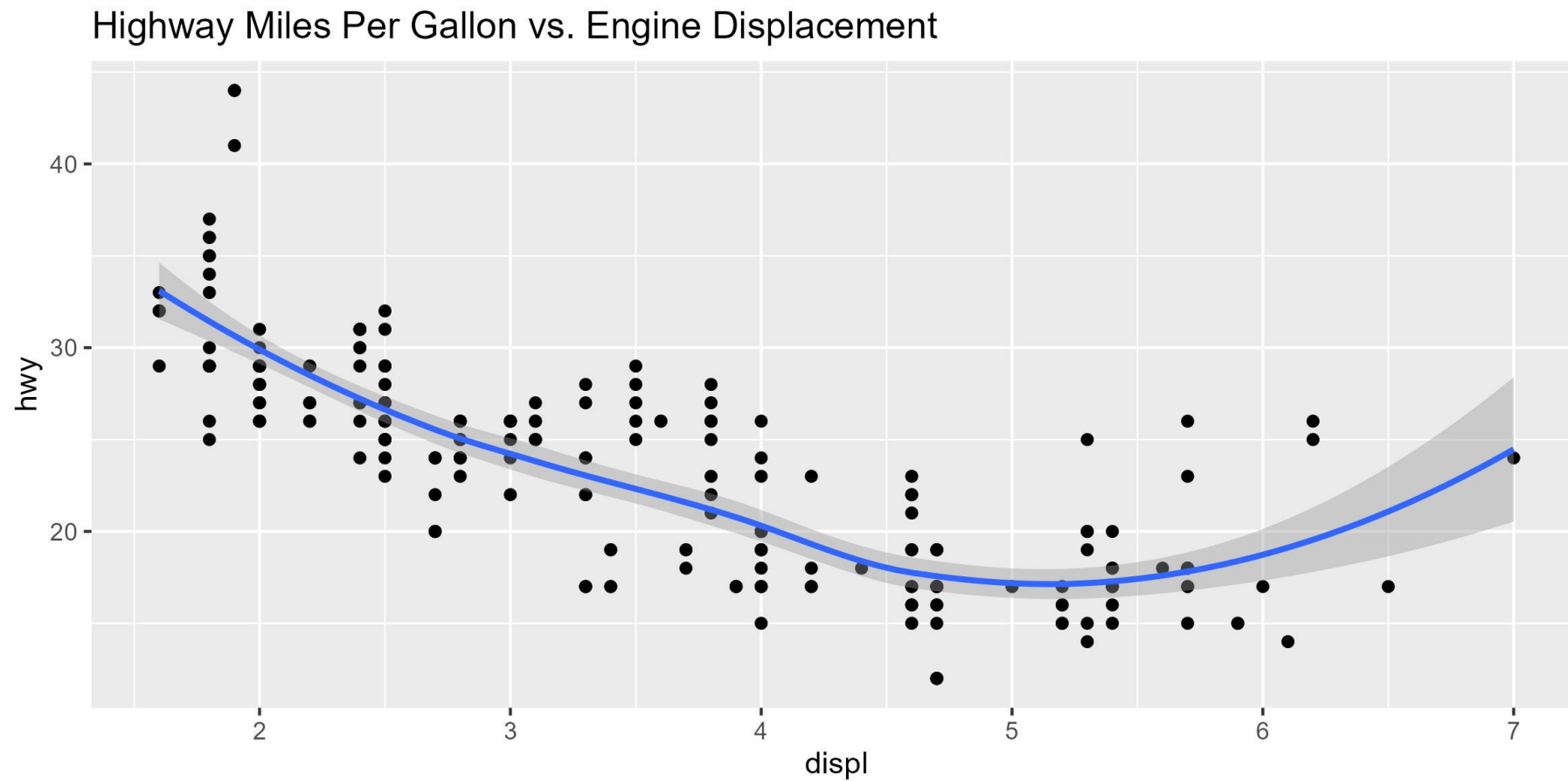




# Design Execution

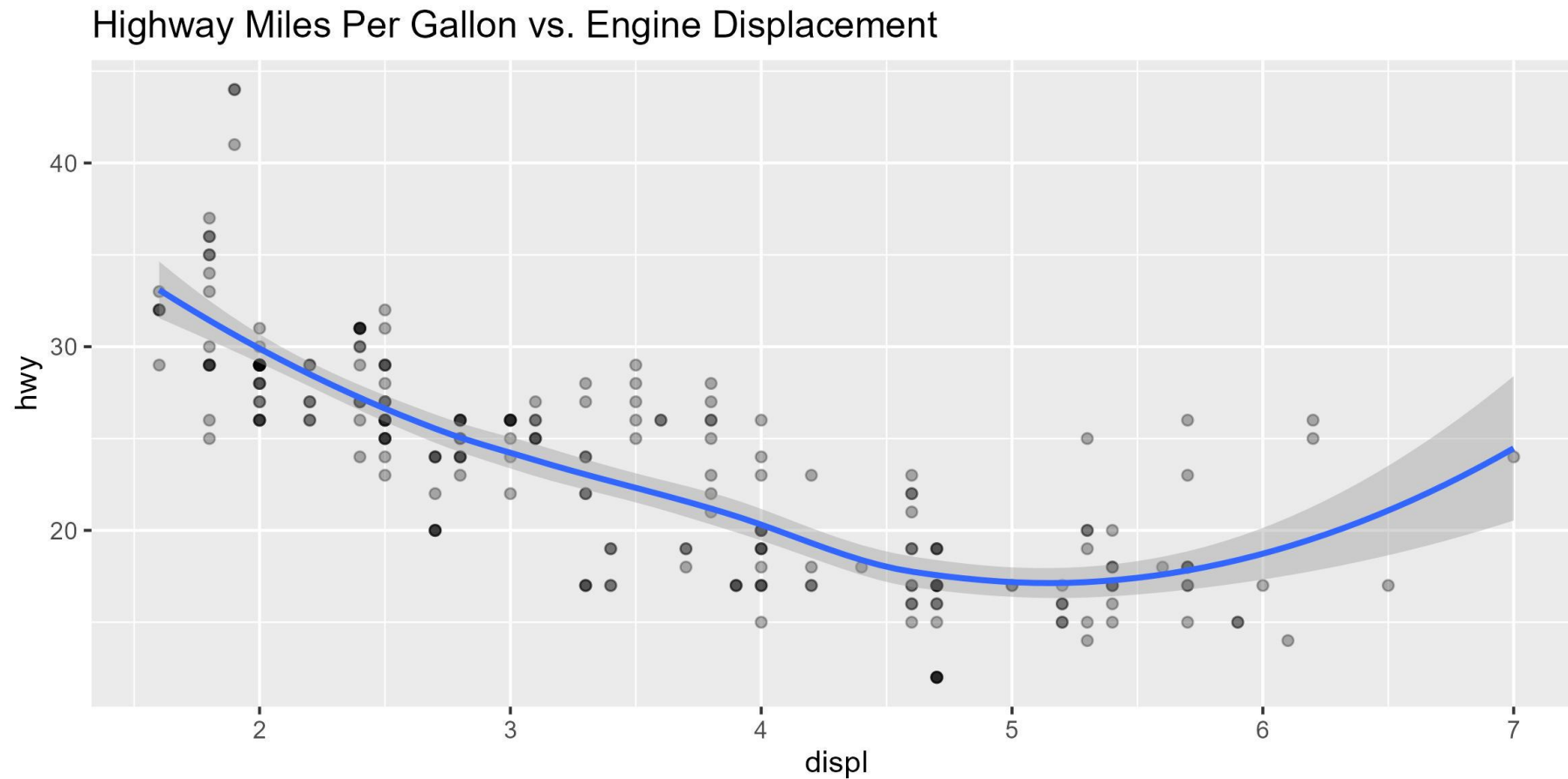
Reproducible R script available on GitHub

# Original



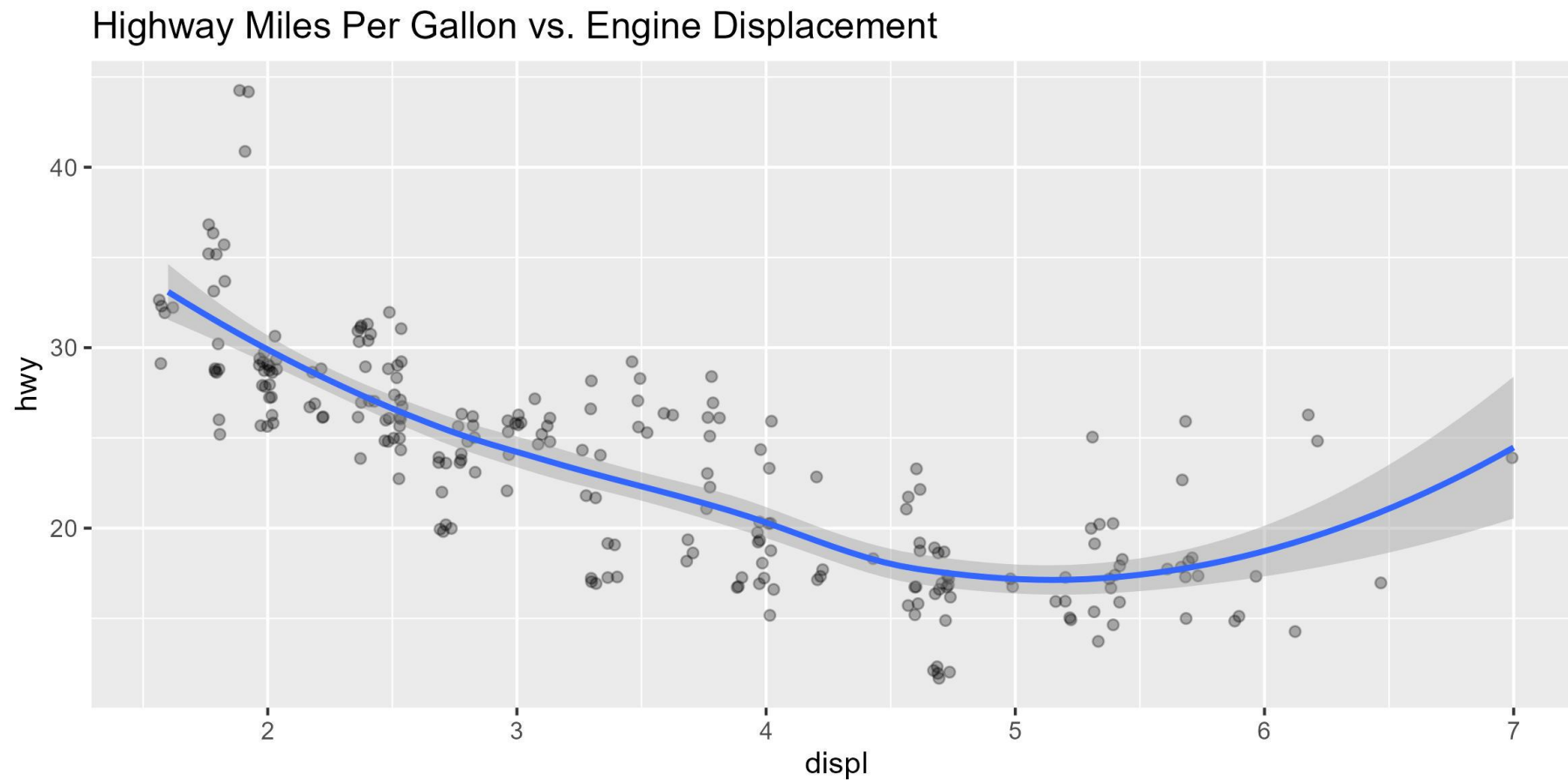
> How would you improve this plot?

# Intentional aesthetics - transparency



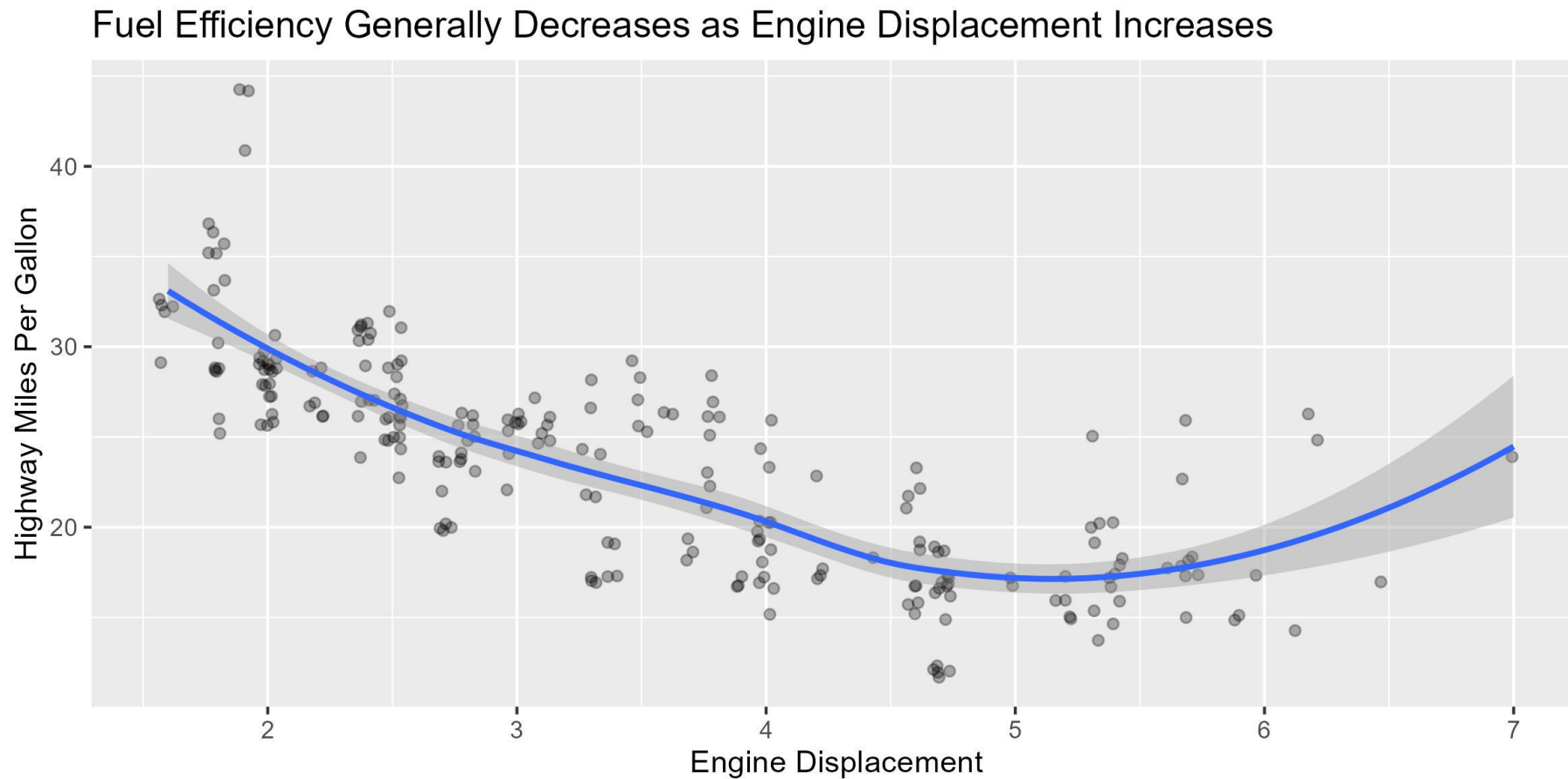
> “...above all else, show the data”

# Intentional aesthetics - position

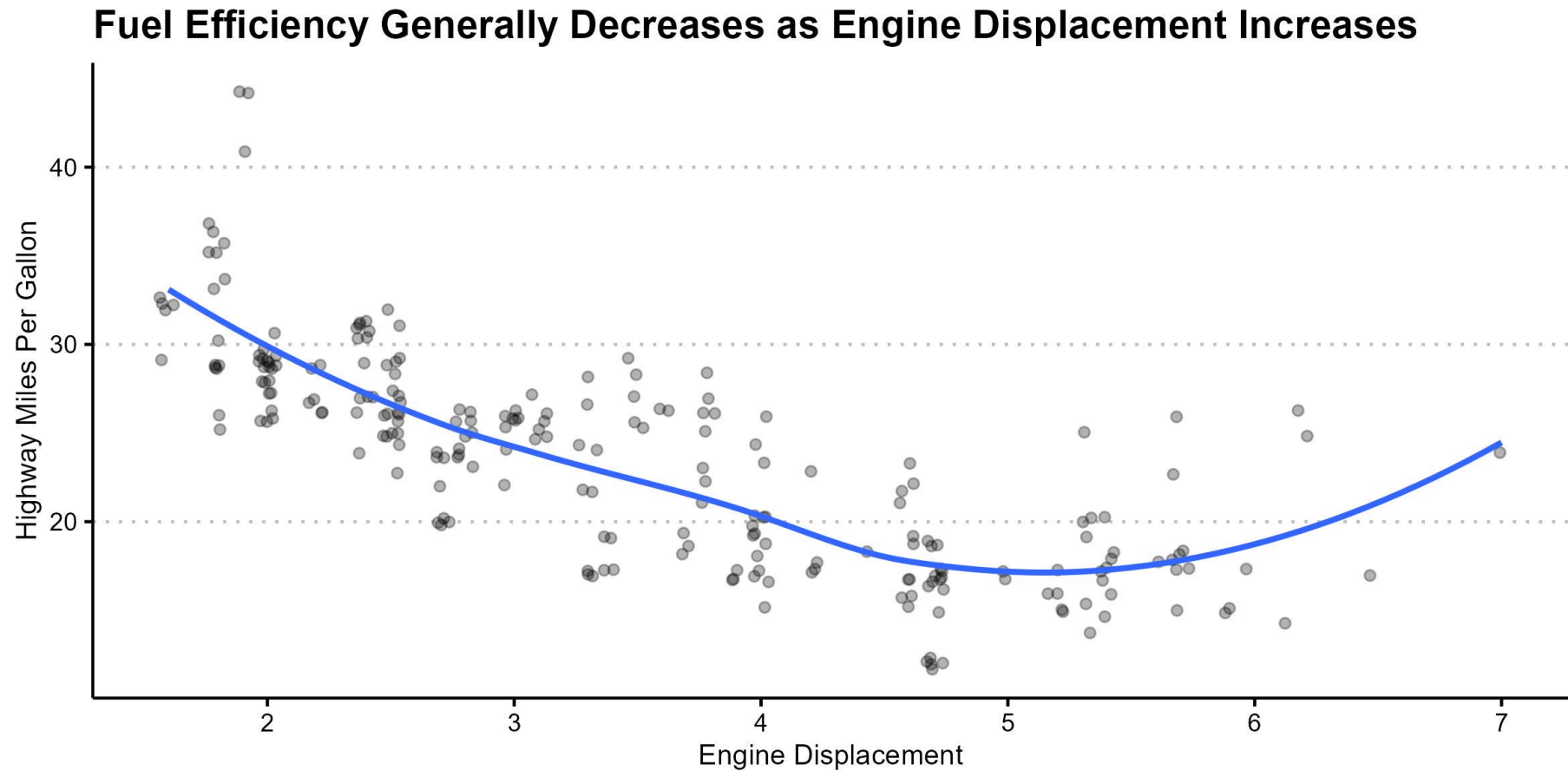




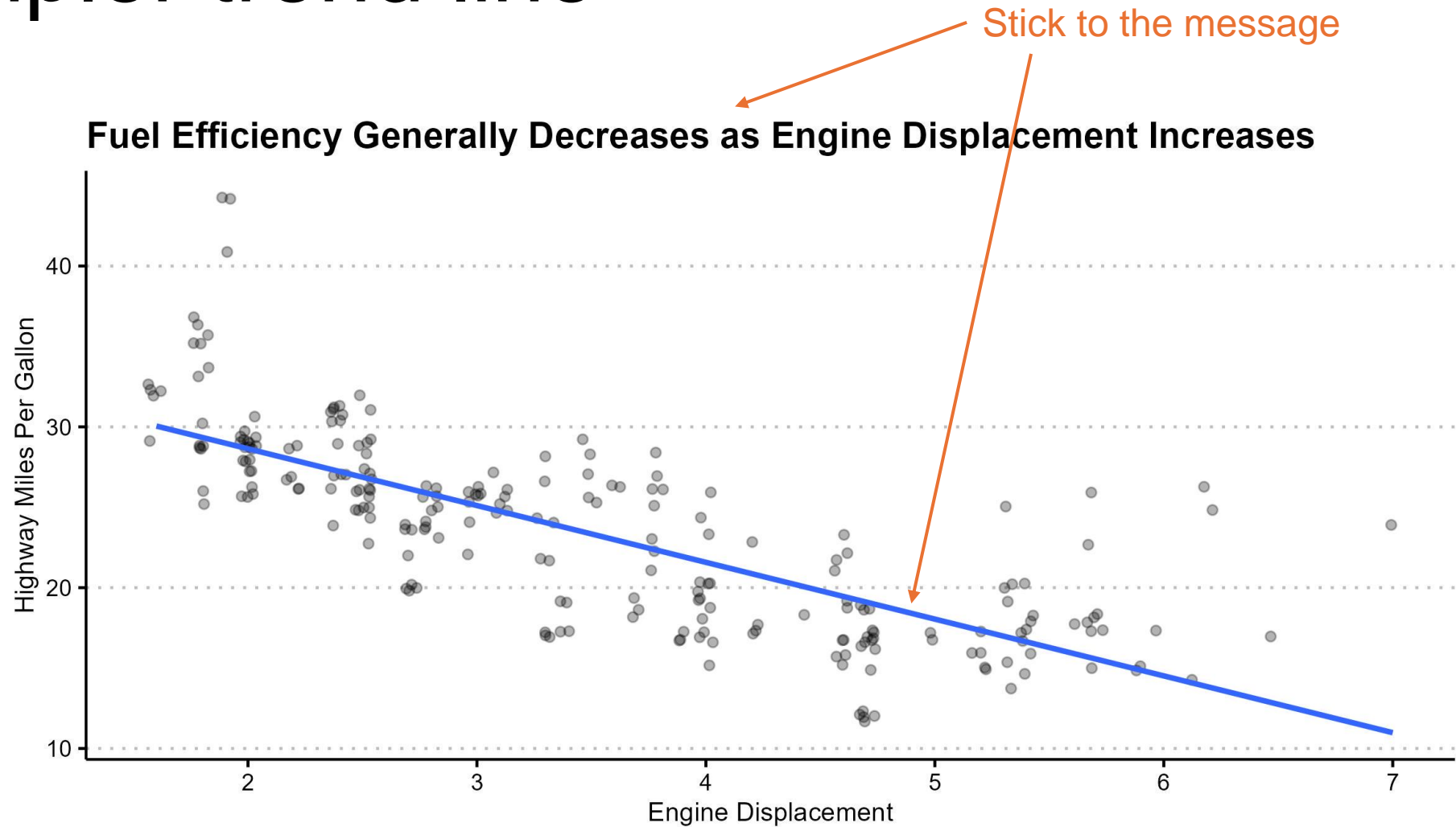
# Descriptive labels and title



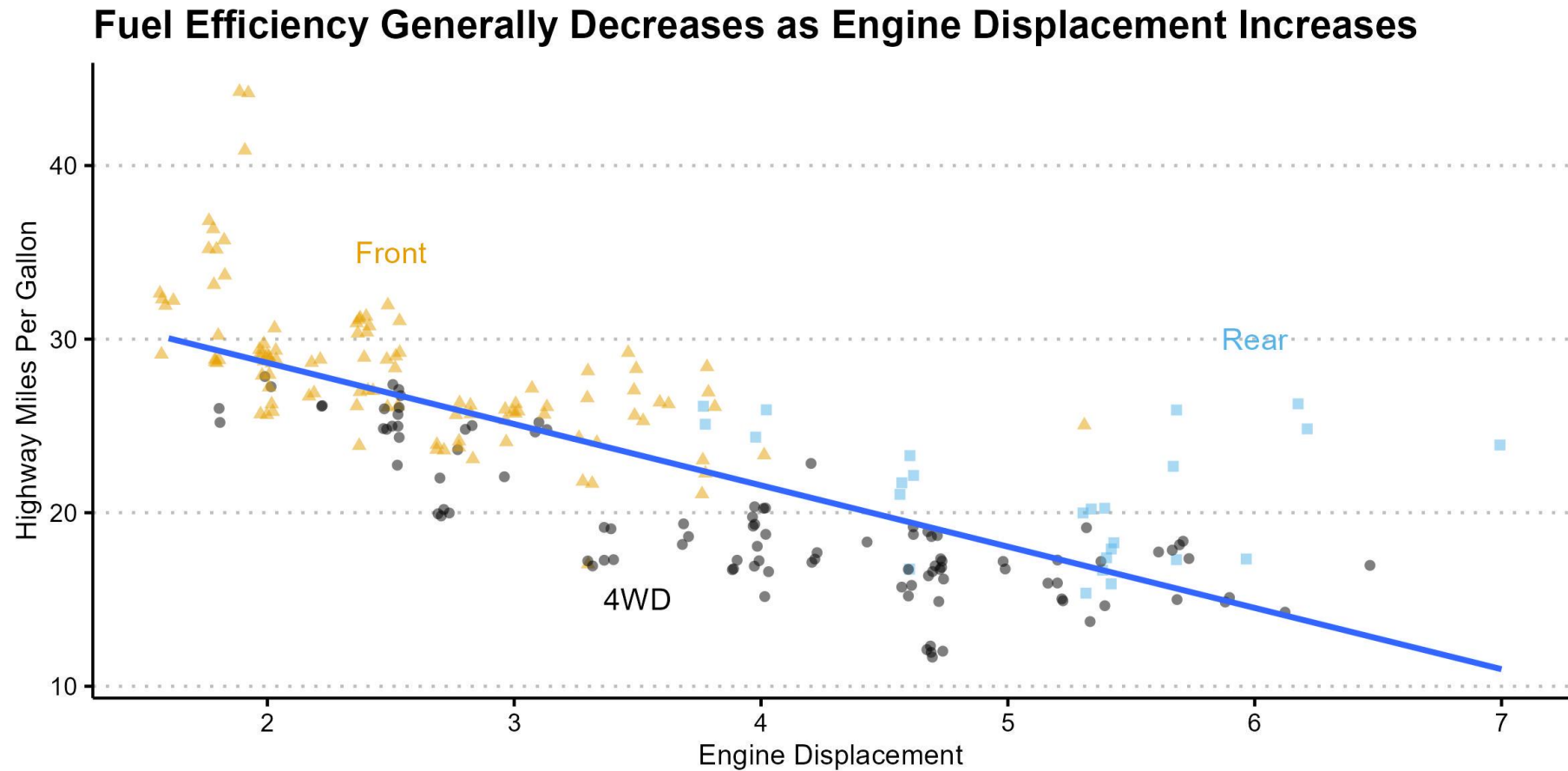
# Remove distractions



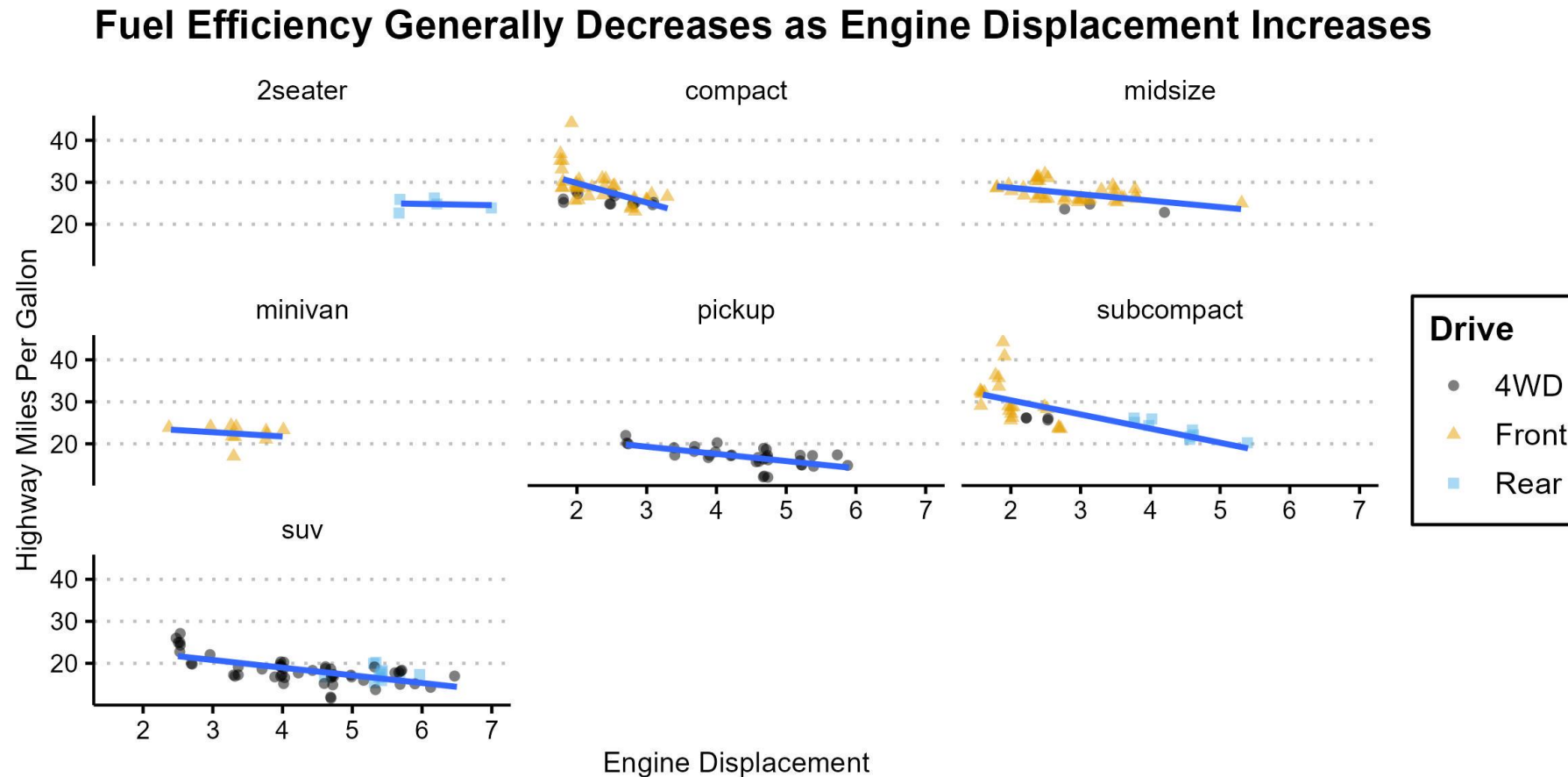
# Simpler trend line



# Increase data density – aesthetics

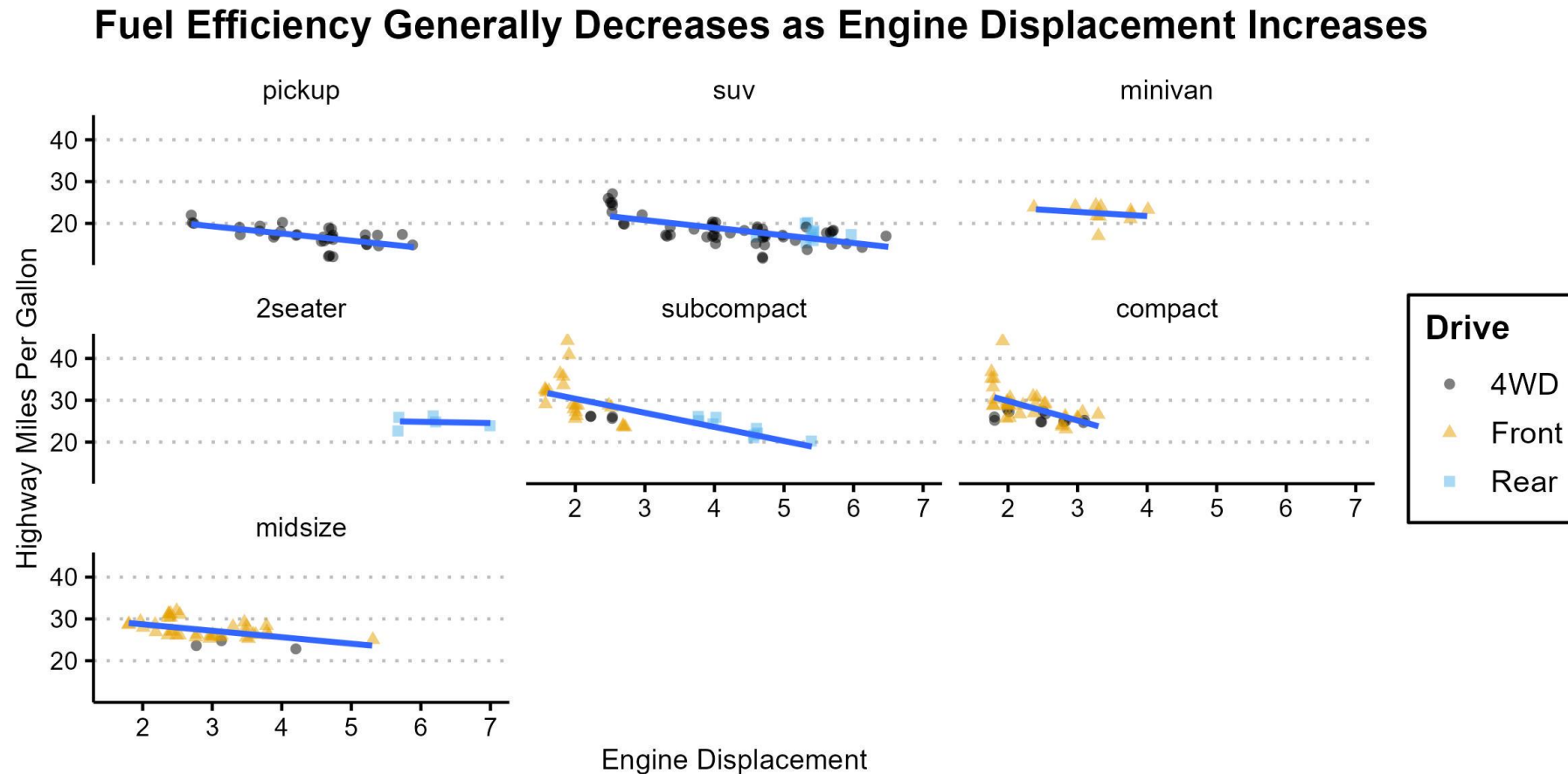


# Increase data density – facets



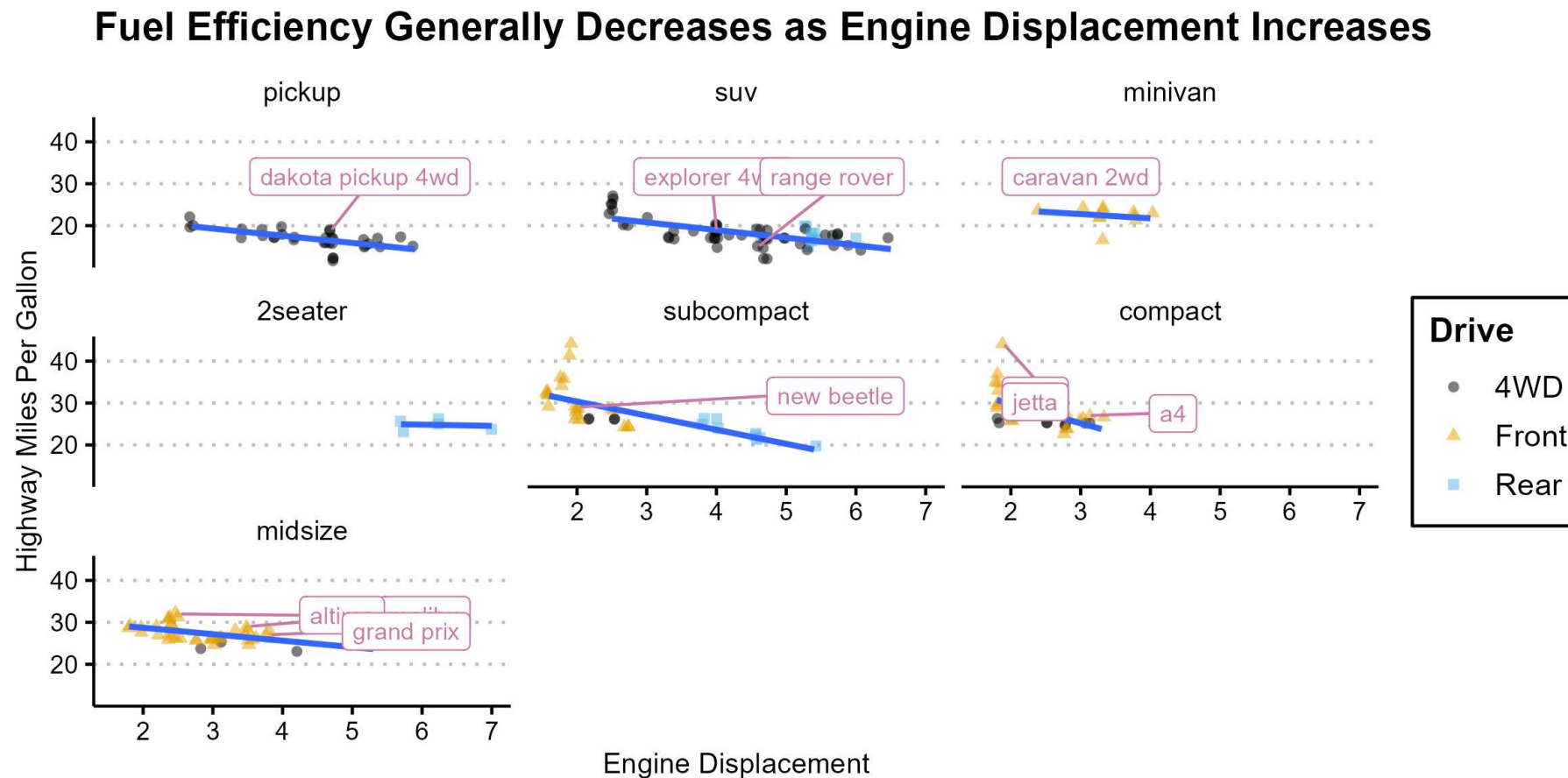
> Why should we not use alphabetical ordering for the panels?

# Increase data density – order matters



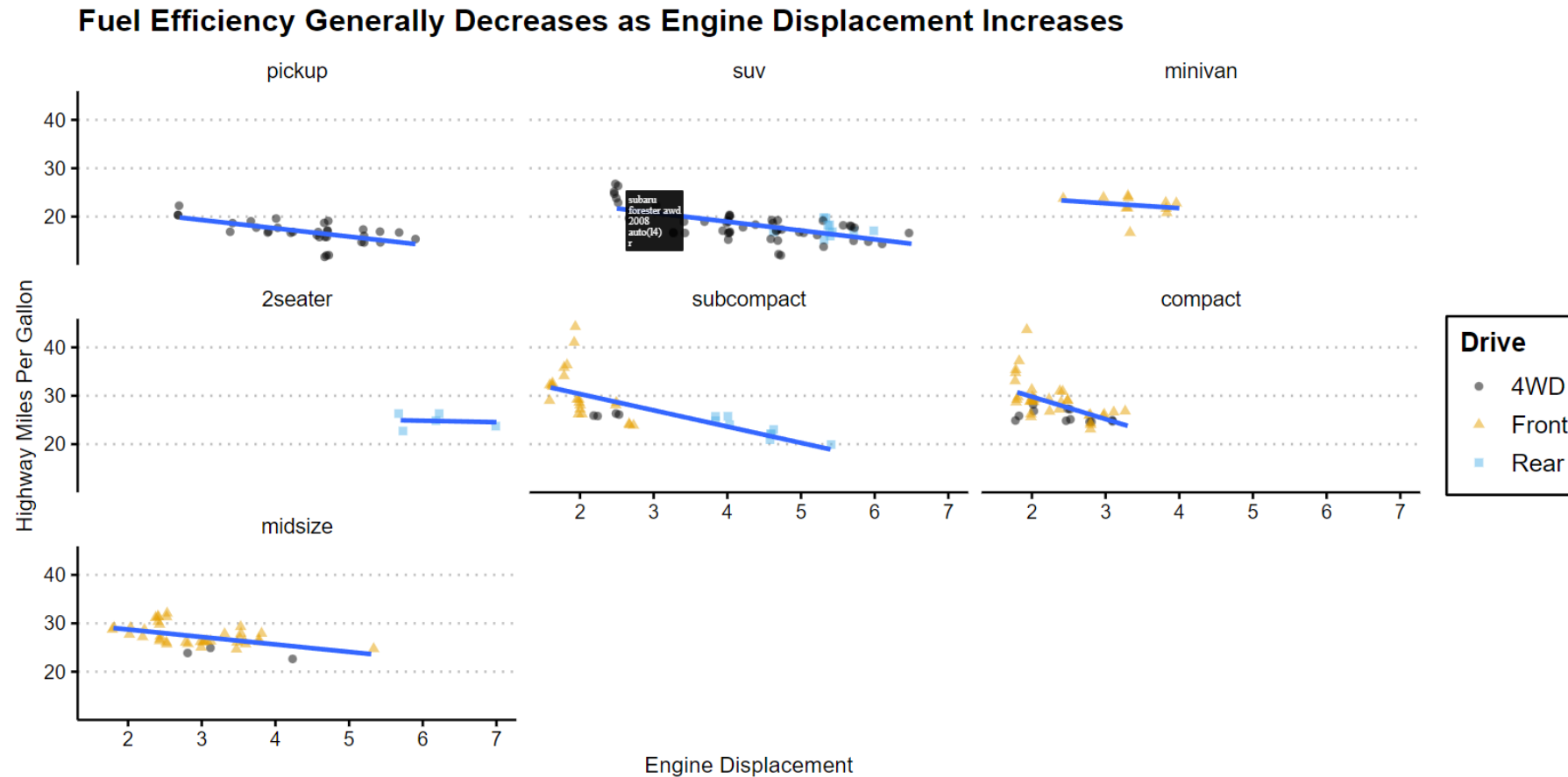
> Order panels by median Highway MPG of each class

# Labels - static



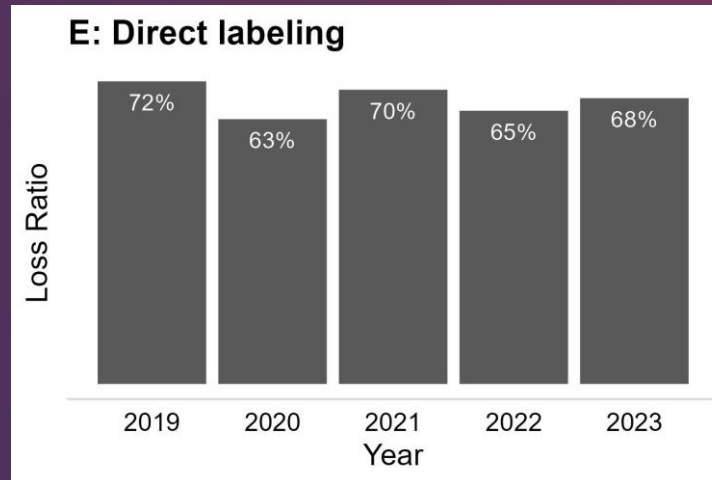
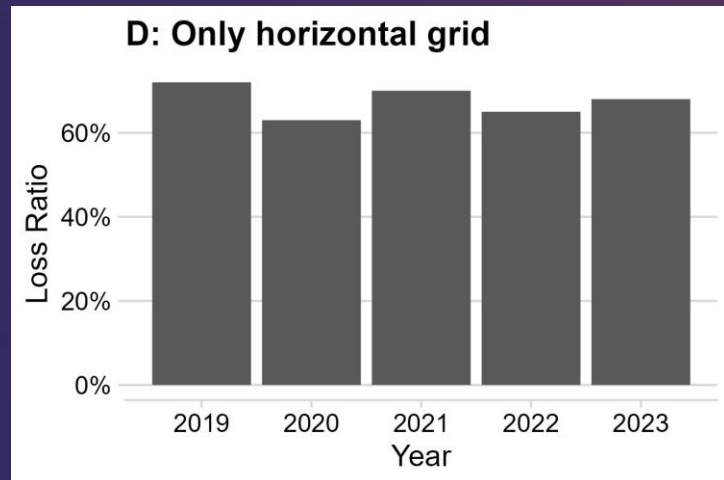
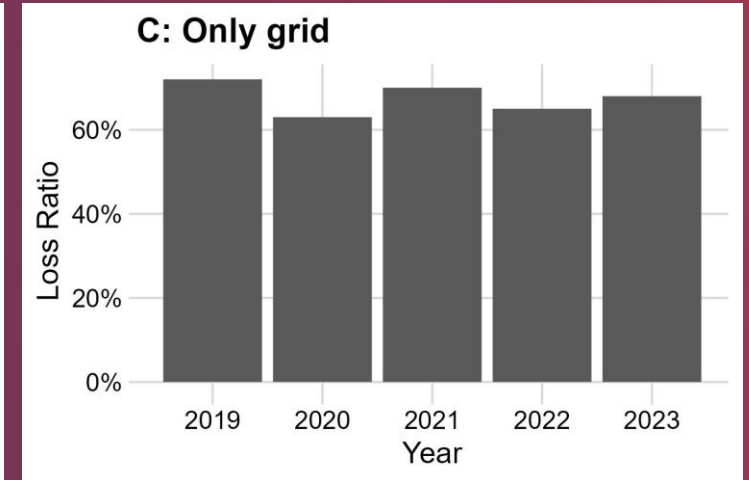
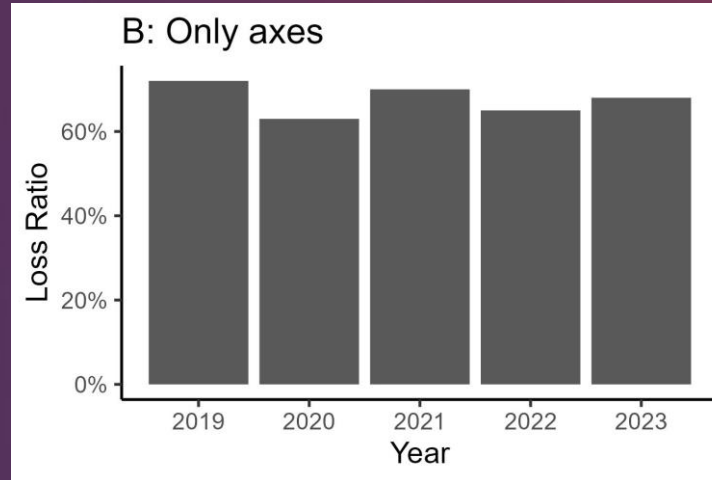
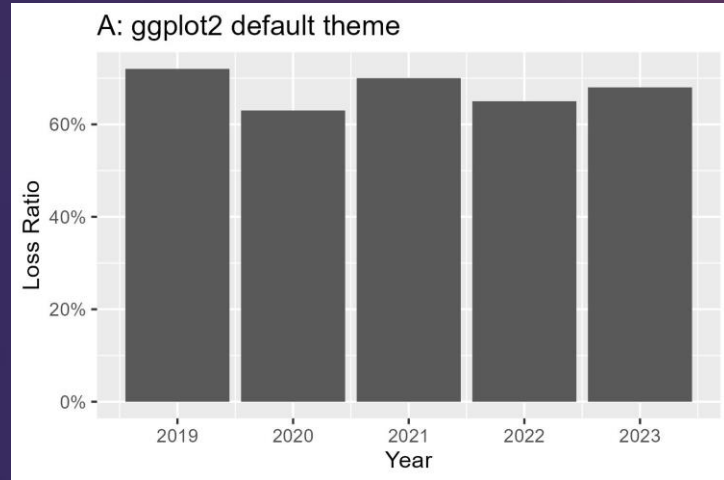
> Helpful but too busy. Anything else we could do?

# Labels - interactive



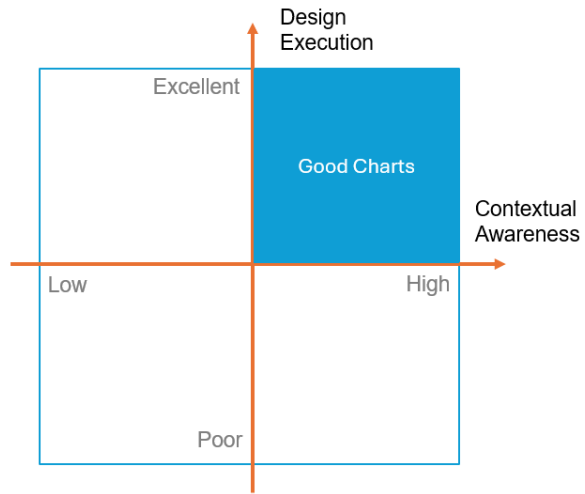


# Poll: Which of the following charts appeals to you the most?



F: Table

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

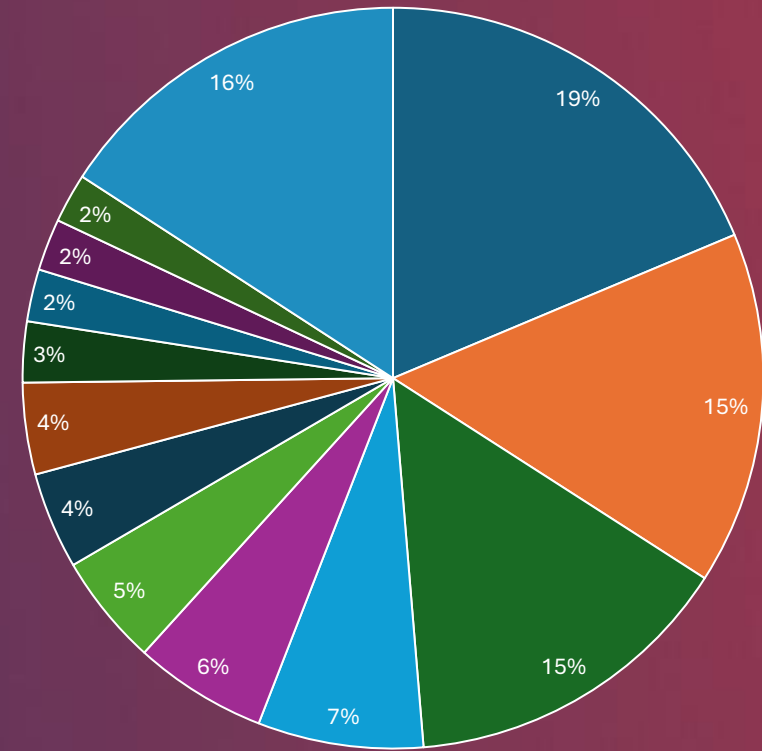
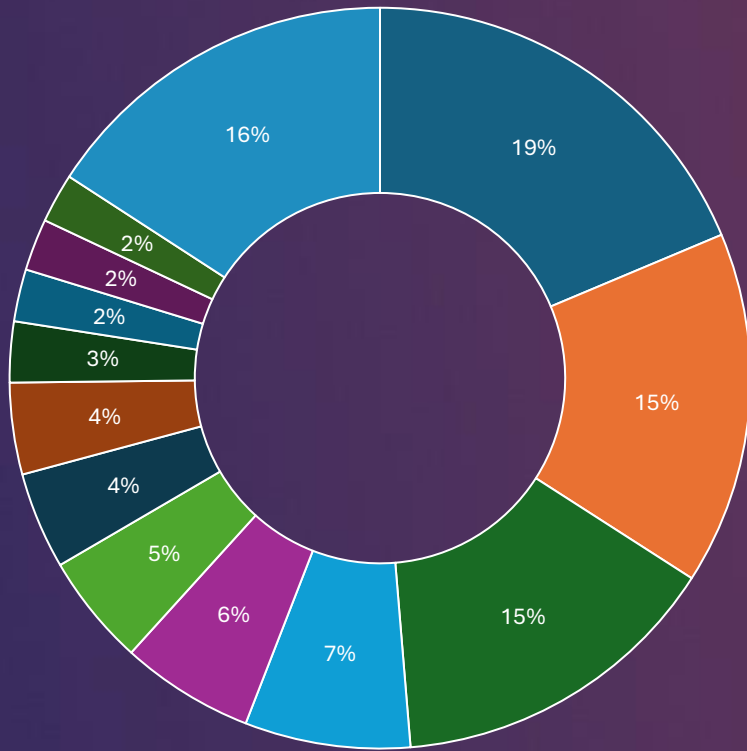


# Contextual Awareness

# Contextual Awareness

What?	Who?	When?	Where?	How?
<ul style="list-style-type: none"><li>• Amounts</li><li>• Distributions</li><li>• Proportions</li><li>• Multivariate relationship</li><li>• Geospatial</li><li>• Uncertainty</li></ul>	<ul style="list-style-type: none"><li>• Actuaries</li><li>• Non-actuaries (UW, brokers, management)</li><li>• Regulators</li></ul>	<ul style="list-style-type: none"><li>• Data exploration phase</li><li>• Results communication phase</li></ul>	<ul style="list-style-type: none"><li>• Board meeting</li><li>• Industry conference</li><li>• Internal team standup call</li></ul>	<ul style="list-style-type: none"><li>• Oral presentation (verbally explained)</li><li>• Written report (textually explained)</li><li>• Infographics (self-explanatory)</li></ul>

# 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



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*

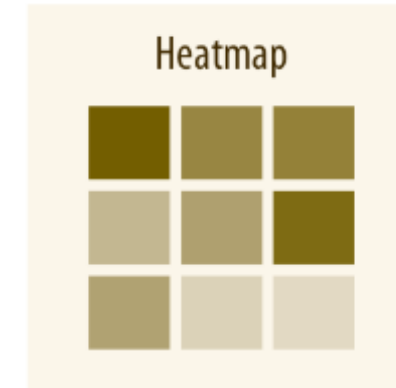
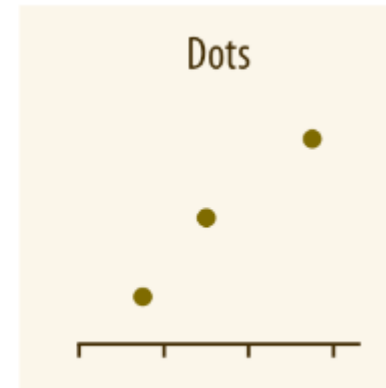
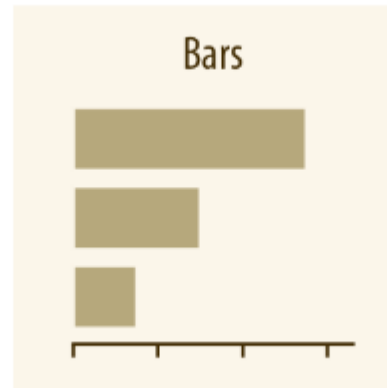
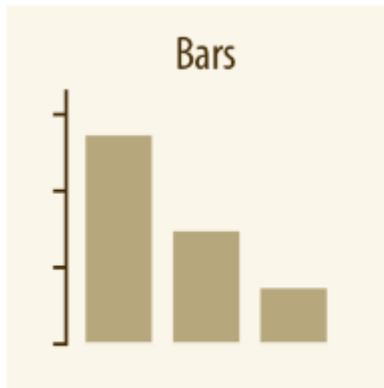
# The “what”



## 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



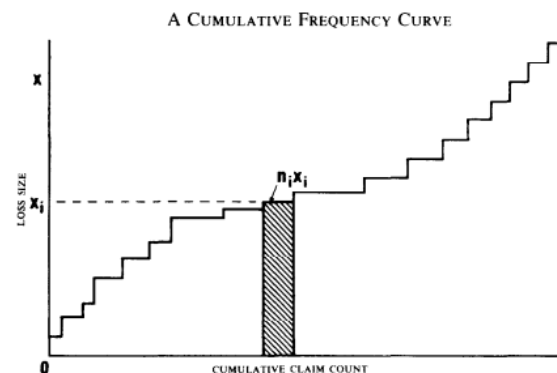
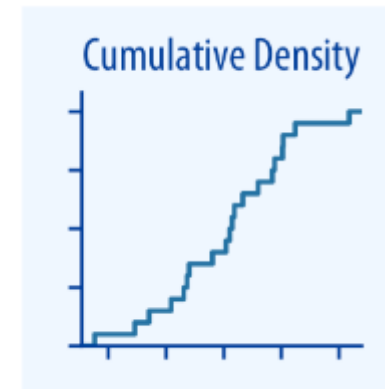
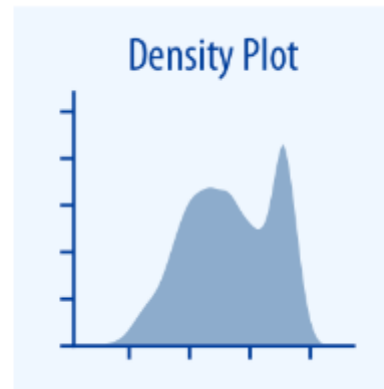
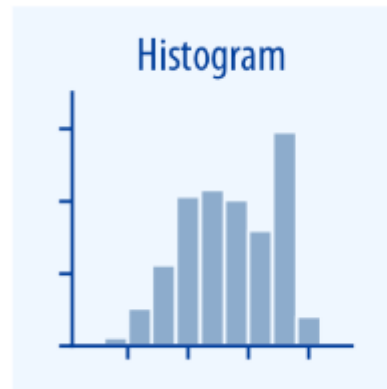
# The “what”



## 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.

- Single distribution



Fundamentals of Data Visualization ([clauswilke.com](http://clauswilke.com))

THE MATHEMATICS OF EXCESS OF LOSS COVERAGES AND  
RETROSPECTIVE RATING-A GRAPHICAL APPROACH ([casact.org](http://casact.org))



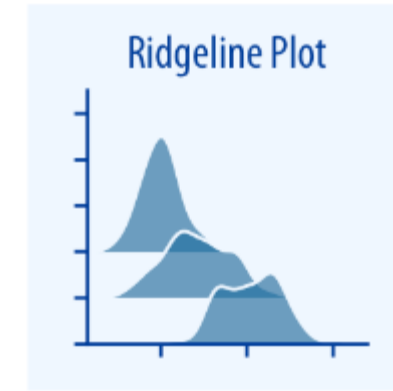
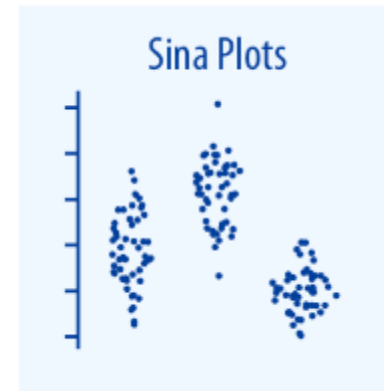
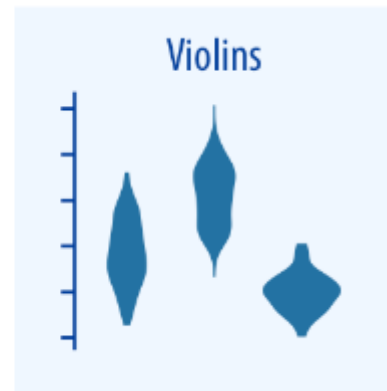
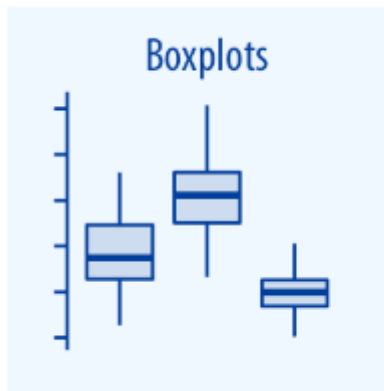
# The “what”



## 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



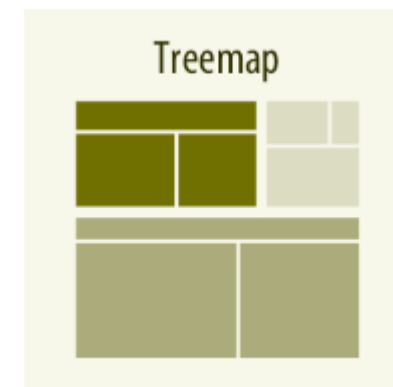
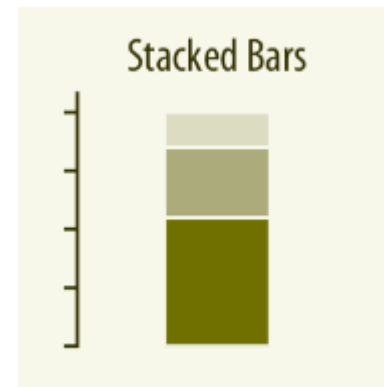
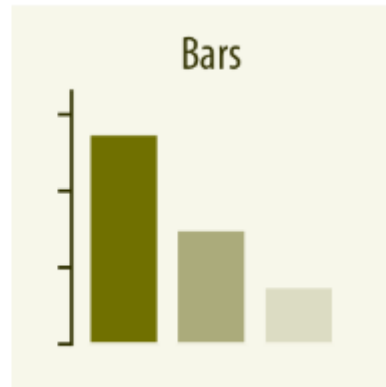
# The “what”



## 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



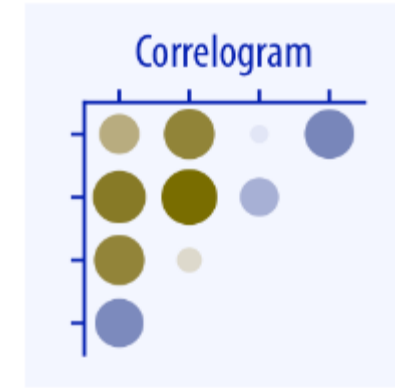
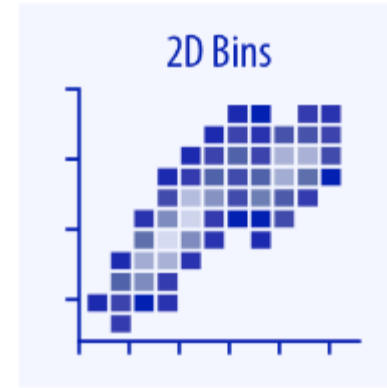
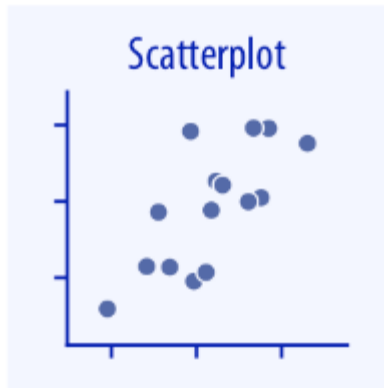
# The “what”



## 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



# 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

Map



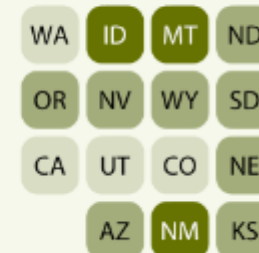
Choropleth



Cartogram



Cartogram Heatmap



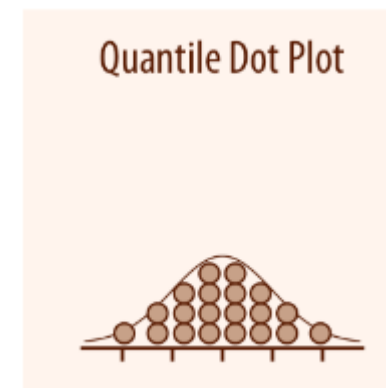
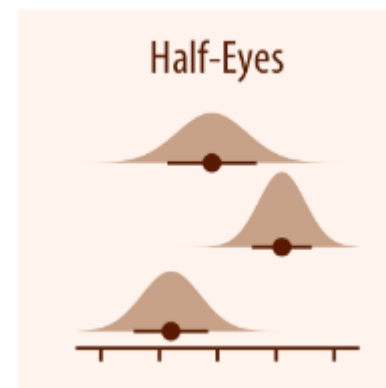
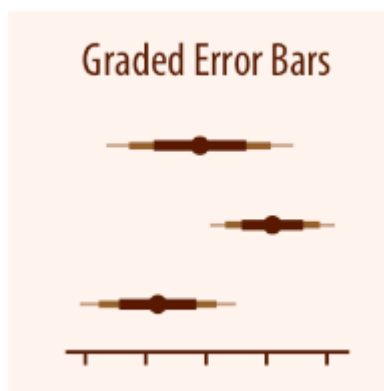
# The “what”



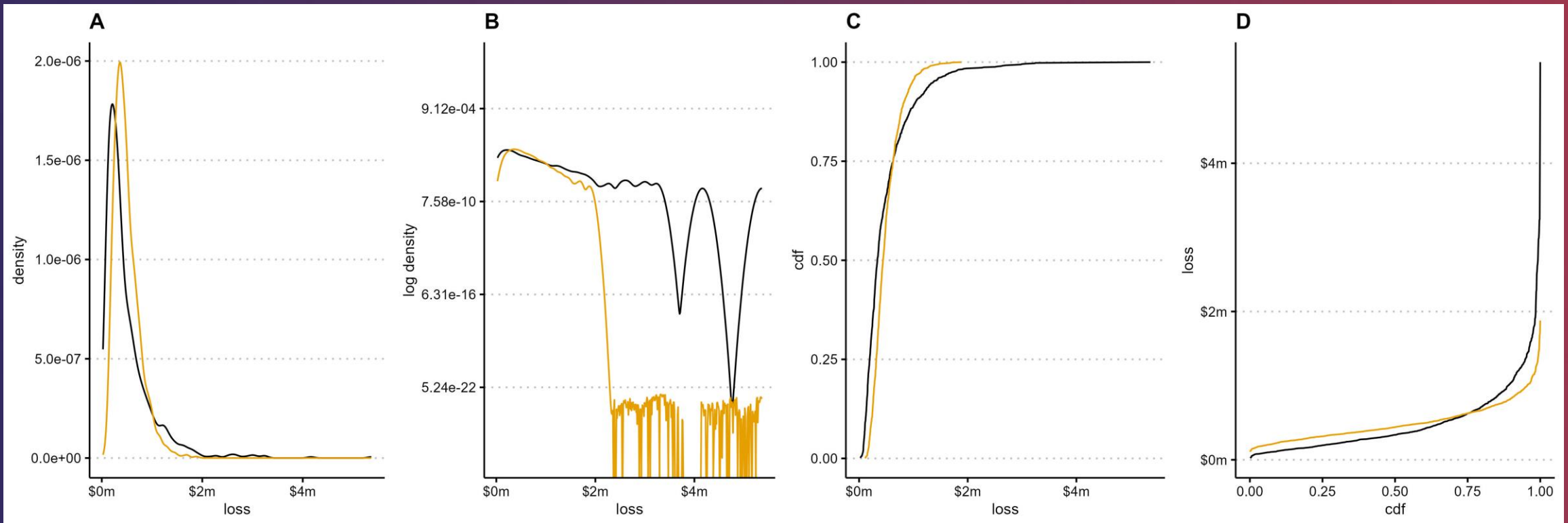
## Practice thoughtfulness:

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.

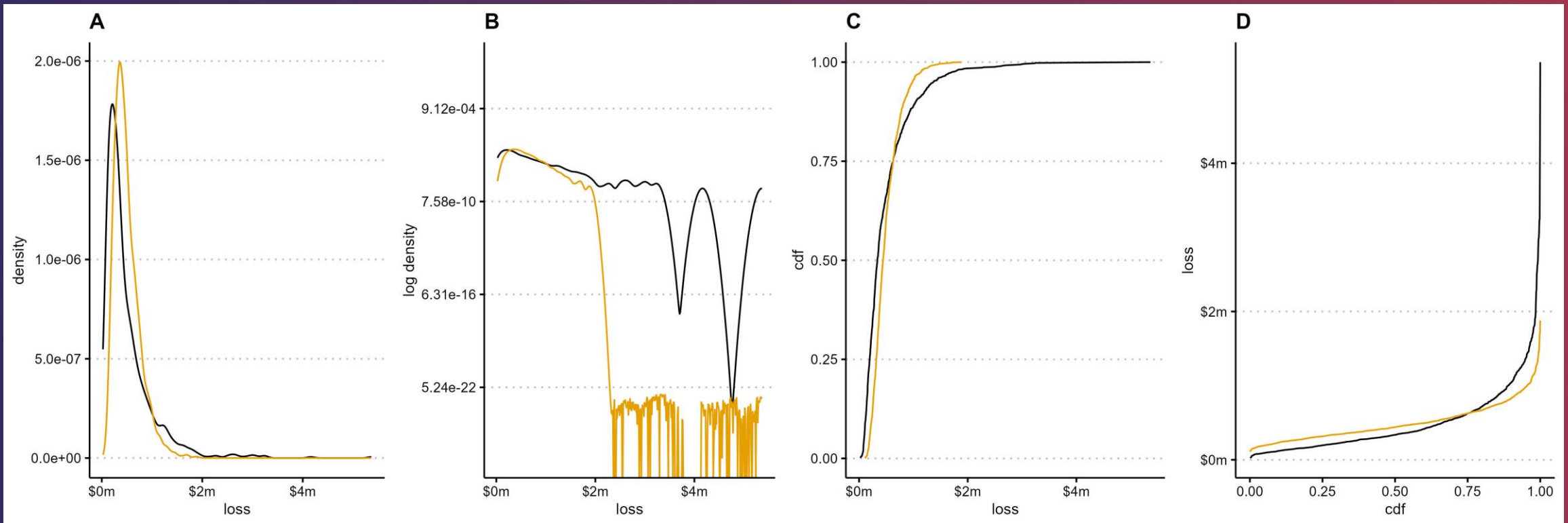
- Uncertainty



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?

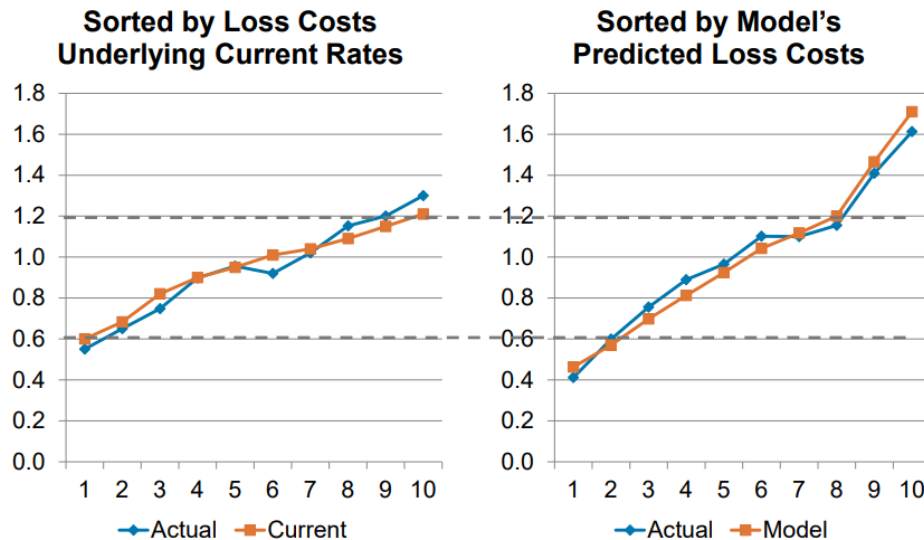


Poll: You want to communicate that a distribution is more heavy-tailed than the other. Which plot below would you choose for a non-actuarial audience?



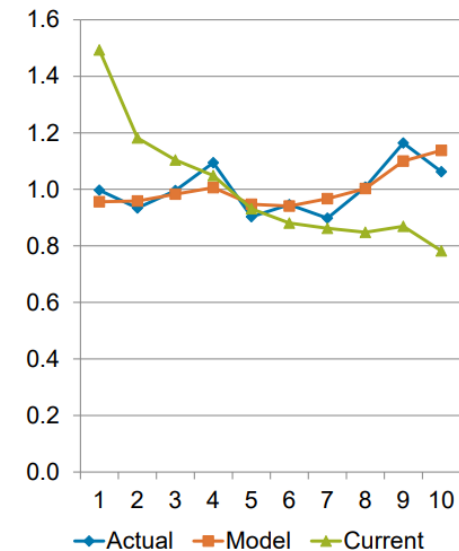
Poll: You want to communicate to management that your new predictive model is superior at differentiating high-cost policyholders from the low-cost ones. Which plot would you choose?

### 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.





# Conclusion

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

# Thank you!

 [enbo.jiang@lockton.com](mailto:enbo.jiang@lockton.com)

 <https://www.linkedin.com/in/enbojiang/>

 <https://github.com/enbojiang/cas-talks>

**Remove**  
to improve  
(the **data-ink** ratio)

Created by Darkhorse Analytics

[www.darkhorseanalytics.com](http://www.darkhorseanalytics.com)

[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)