CAS

2024 Spring Meeting

ATL | Adapt Transform Lead

May 5-8 2024 • Atlanta, Georgia

1

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

Visuary

[noun]

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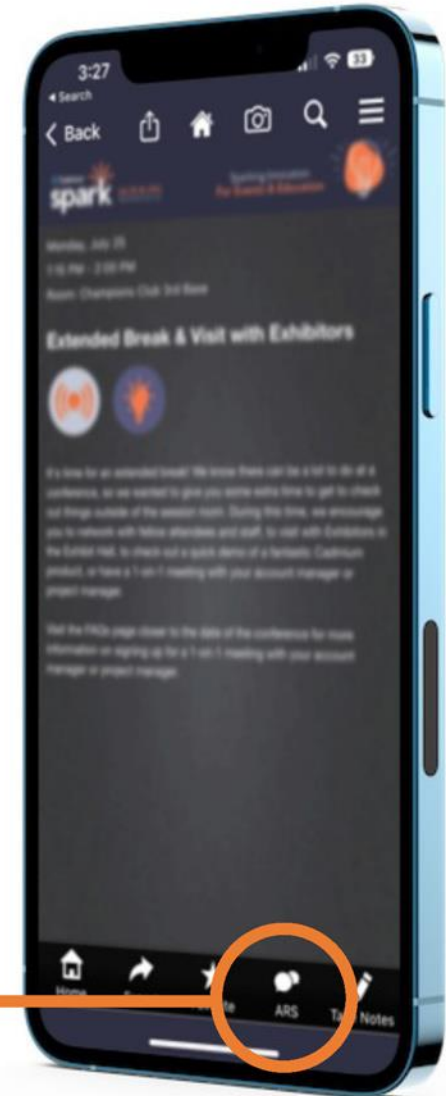
Agenda

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

Get your devices ready for a poll!

1. Locate the session in the mobile app.

2. Tap the " 💬 " button on the bottom menu bar of the session to open the poll.

3. The poll will launch when polling begins.

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?!

This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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 (!)

This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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

Visual Thinking

- Data Visualization for Actuaries and Data Scientists

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?

20XX Net Earned Premium by Line

Or this?

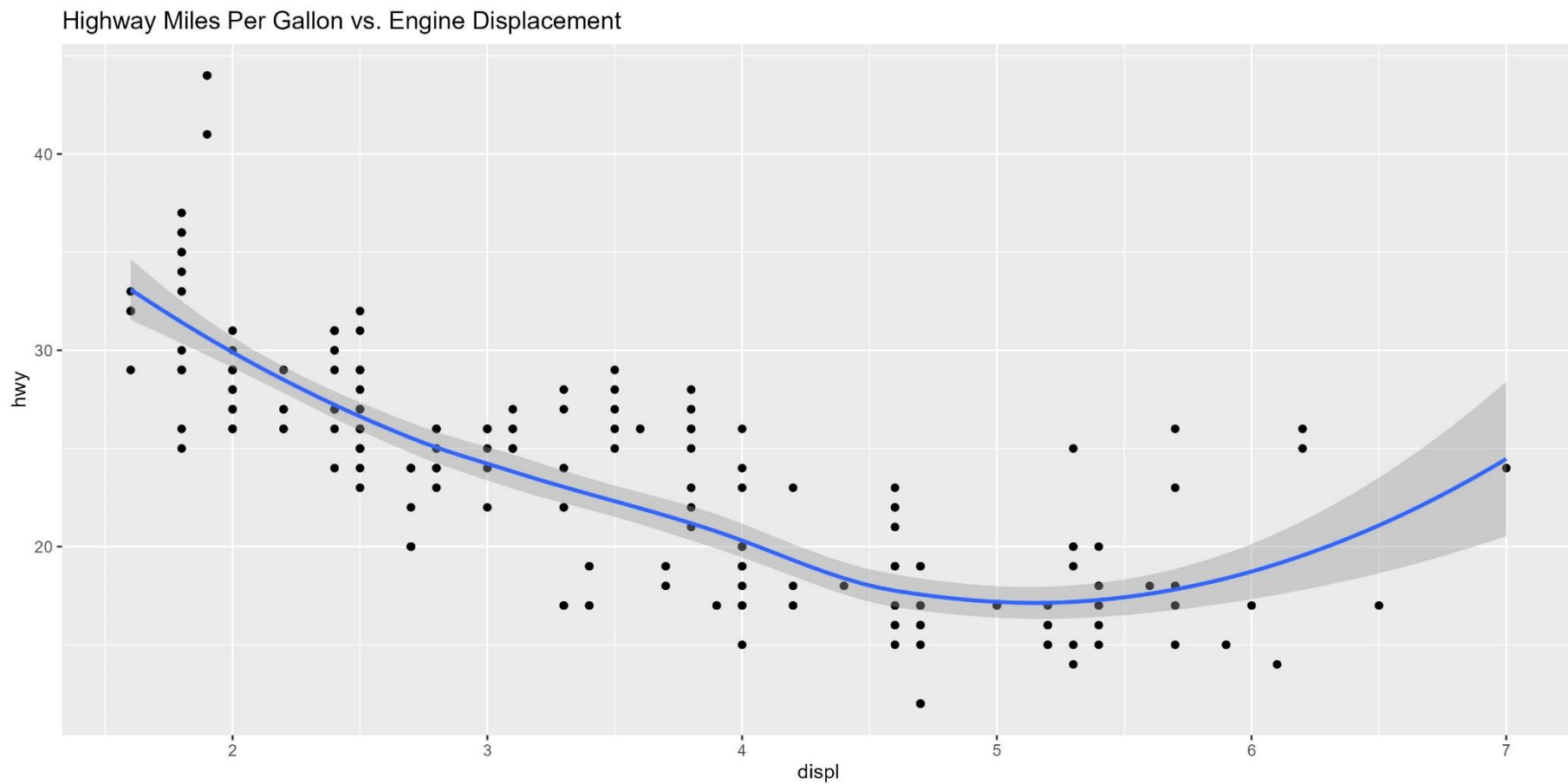# Or this?

Ultimate Nominal Frequency

Ultimate Severity
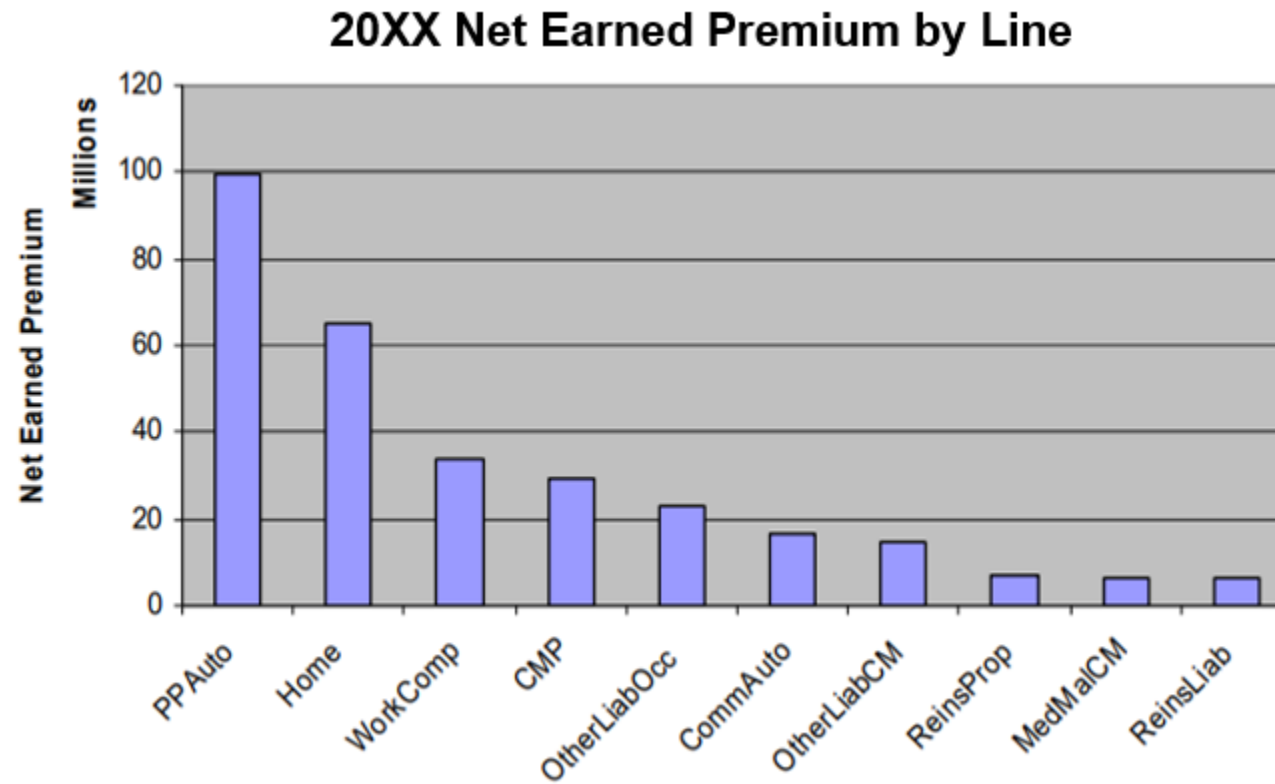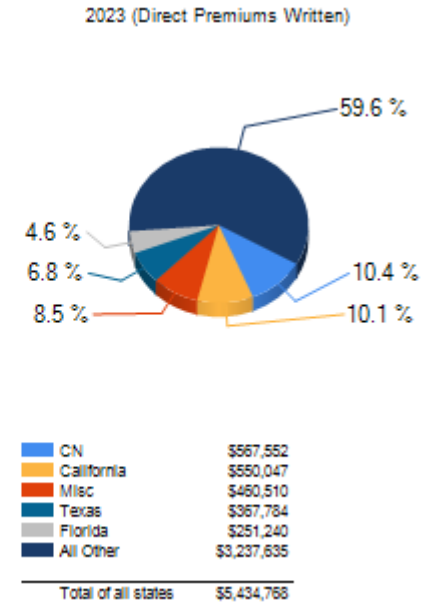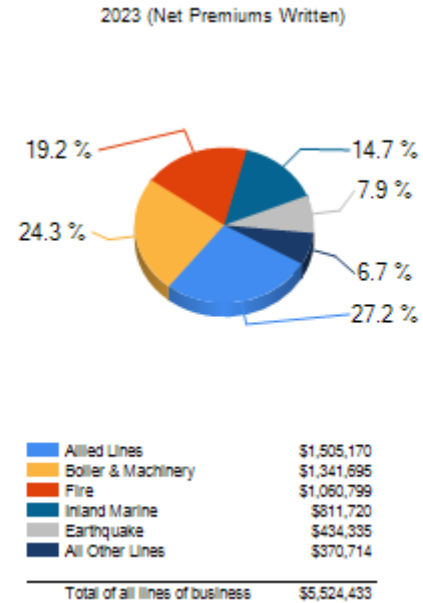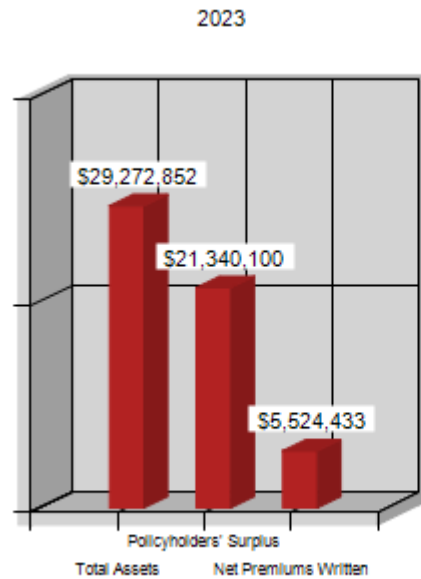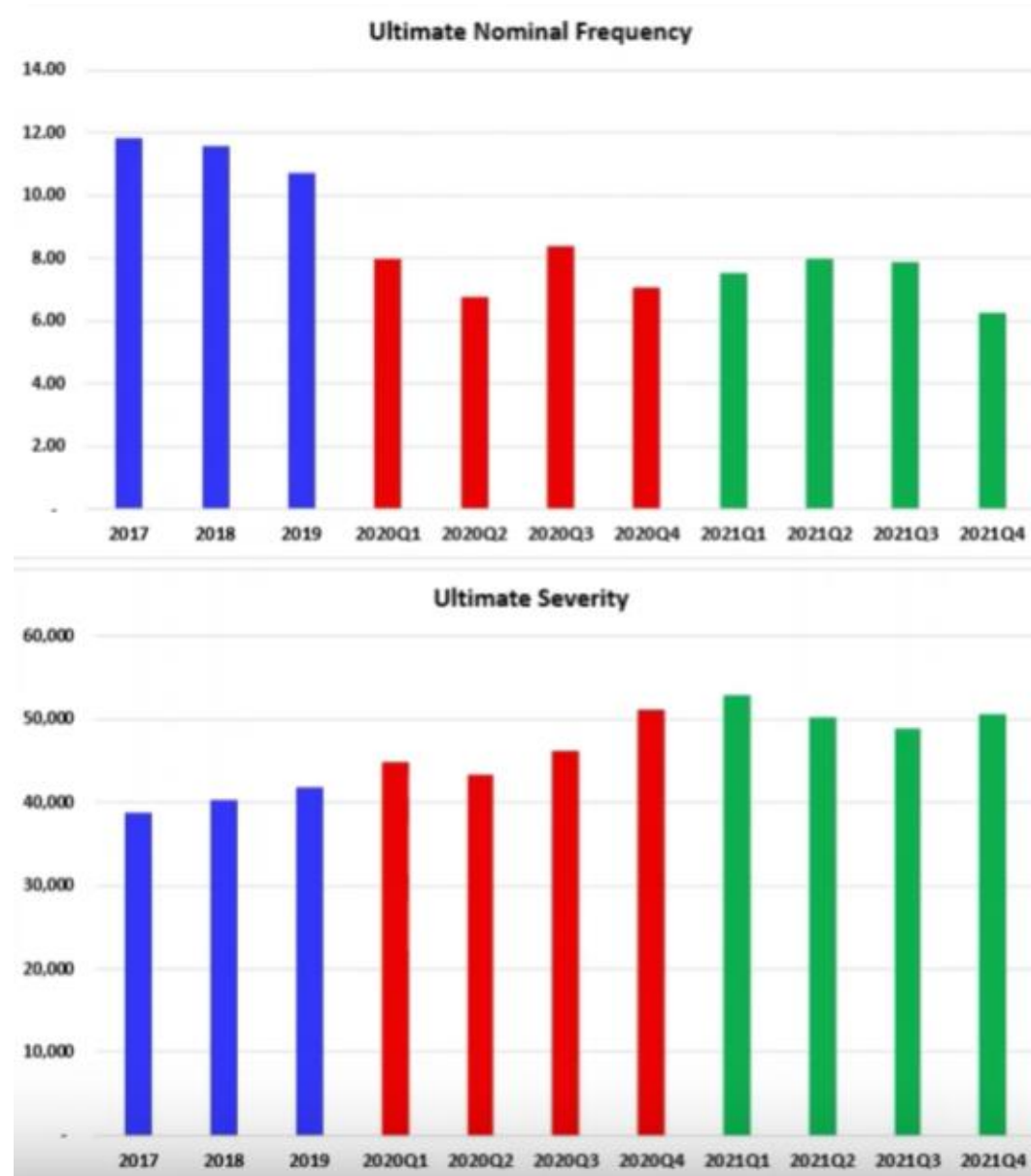
An unnamed industry
benchmark provider

16

What about this?

Highway Miles Per Gallon vs. Engine Displacement
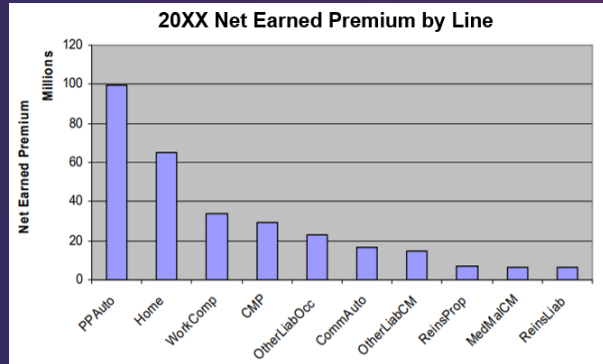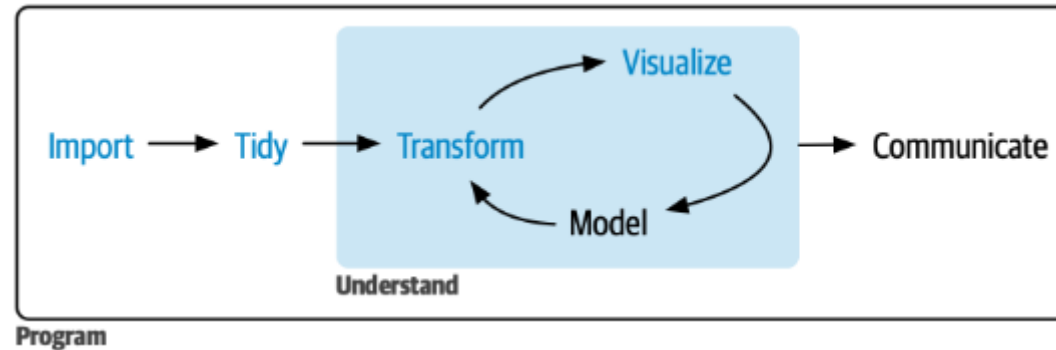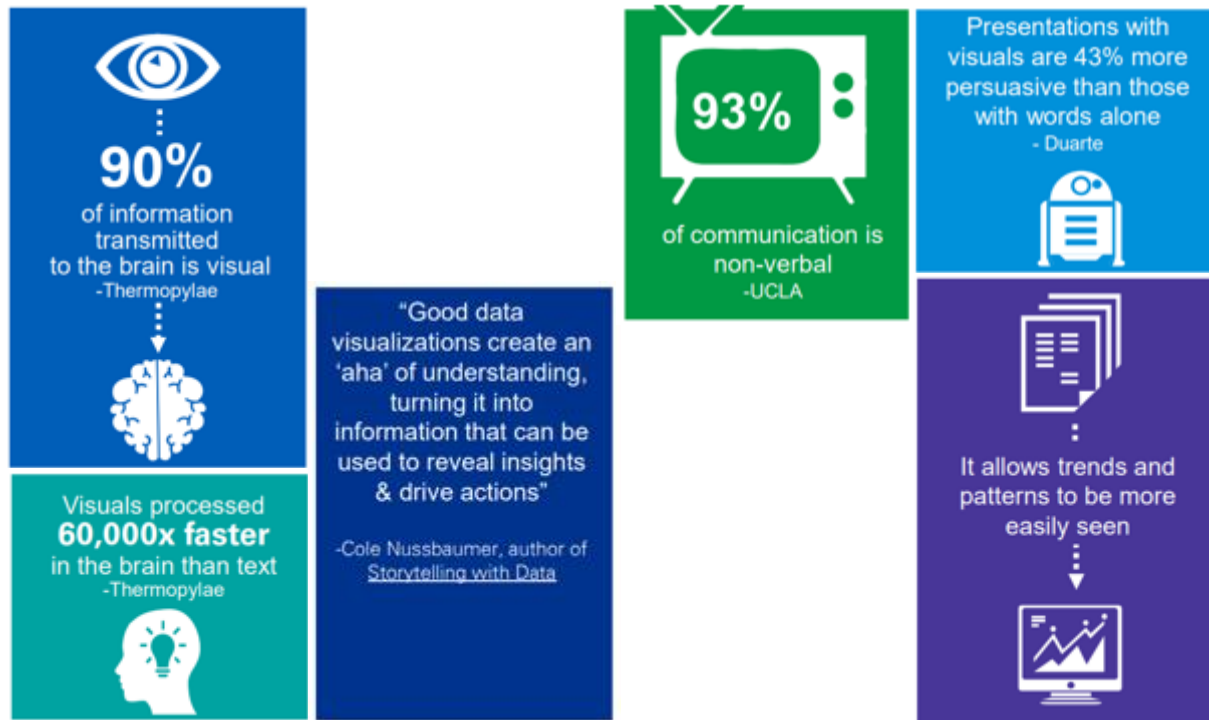
How can we do better?
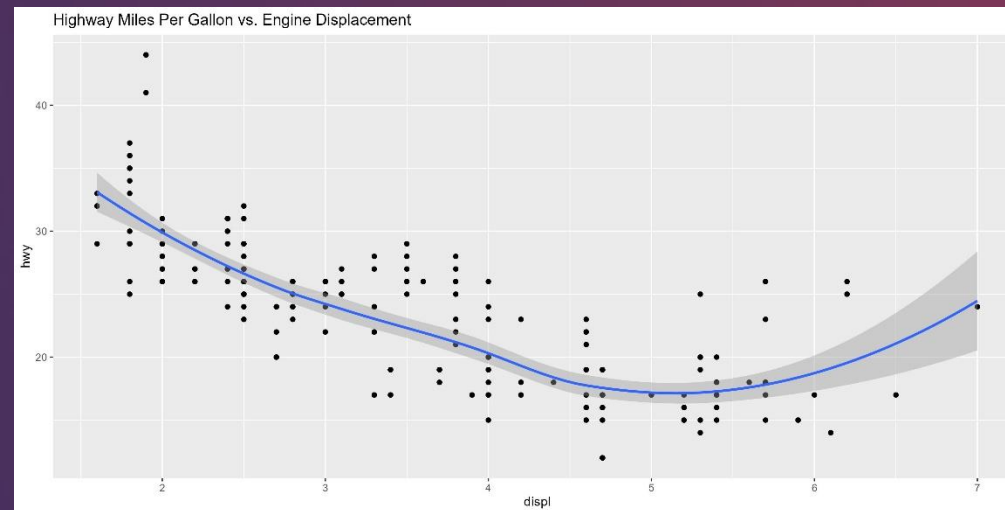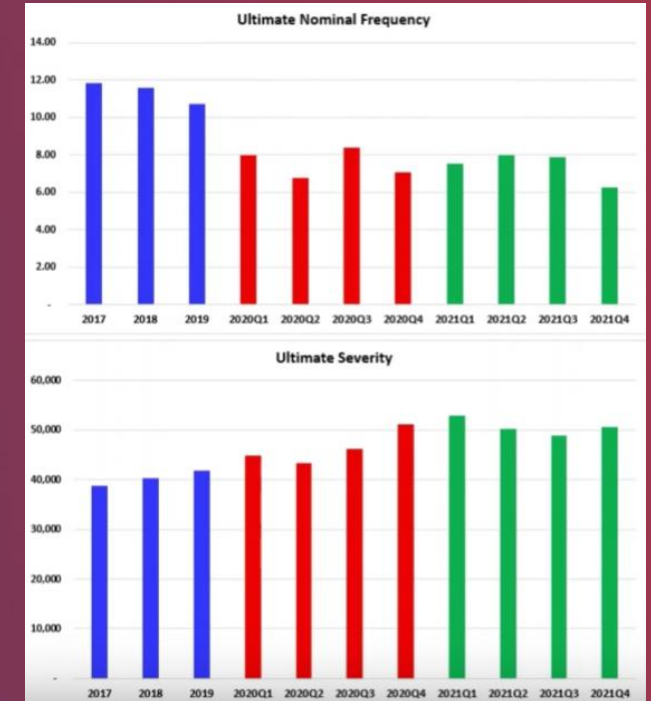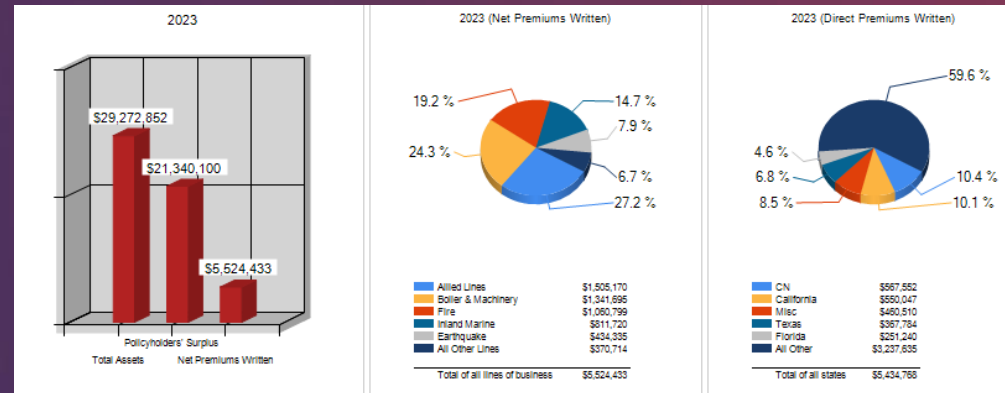
Why should we care?

- 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

90% of information transmitted to the brain is visual -Thermopylae

Visuals processed 60,000x faster in the brain than text -Thermopylae

"Good data visualizations create an 'aha' of understanding, turning it into information that can be used to reveal insights & drive actions" -Cole Nussbaumer, author of Storytelling with Data

93% of communication is non-verbal -UCLA

Presentations with visuals are 43% more persuasive than those with words alone - Duarte

It allows trends and patterns to be more easily seen

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 DISC DA Syllabus (casact.org)

CAS Capability Model

Professionalism
Ability to act in a professional manner and make ethical decisions aligned with professional actuarial standards.

Data Analysis
Skills to gather, analyze, and draw practical conclusions from data as well as communicate data findings to others.

Business and Technical Communication
Ability to share information about professional topics with individuals inside and outside of one's own organization to drive business outcomes.

Critical Thinking & Problem Solving
Skills to think through a problem by evaluating each component part and applying prior knowledge to find one or more solutions.

Business Acumen
Ability to apply knowledge of insurance products and business problems to make decisions.

Leveraging Technology
Ability to use technology to improve processes and outcomes.

Level 1:
Reviews and evaluates data (e.g., data cleaning, confirm integrity), manipulates data through aggregation and exploration, and documents findings.

Level 2:
Uses advanced data visualization and statistical tools, programming, and machine learning tools to anticipate and address competing explanations, validate results, and provide insights.

Level 3:
Works with complex data, across data sources, sets data governance standards, and provides business critical insights.

Traits

Content Areas

Skills

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 <u>not</u>

- 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 <mark>**greatest number of ideas** **in the shortest time** with the least ink **in the smallest space**</mark>
  - 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"

Visuaries do not mislead

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

Obvious
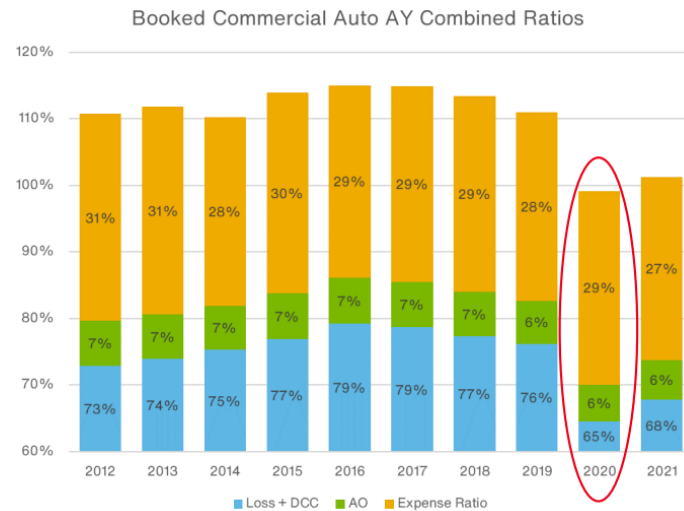
Booked Commercial Auto AY Combined Ratios

Subtle

It's Who You Know

By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these **personal social networks** begin to decline over time.

Interacting with other entrepreneurs fosters possible role models, networking opportunities, advice, and encouragement.
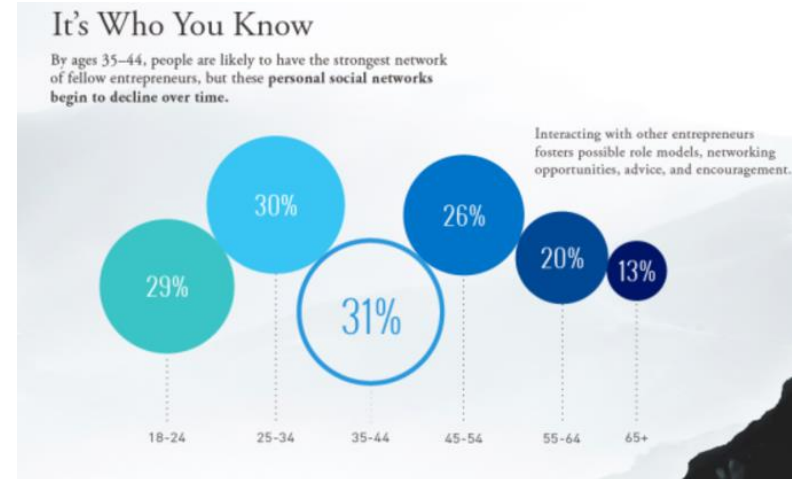
Visuaries employ intentional aesthetics
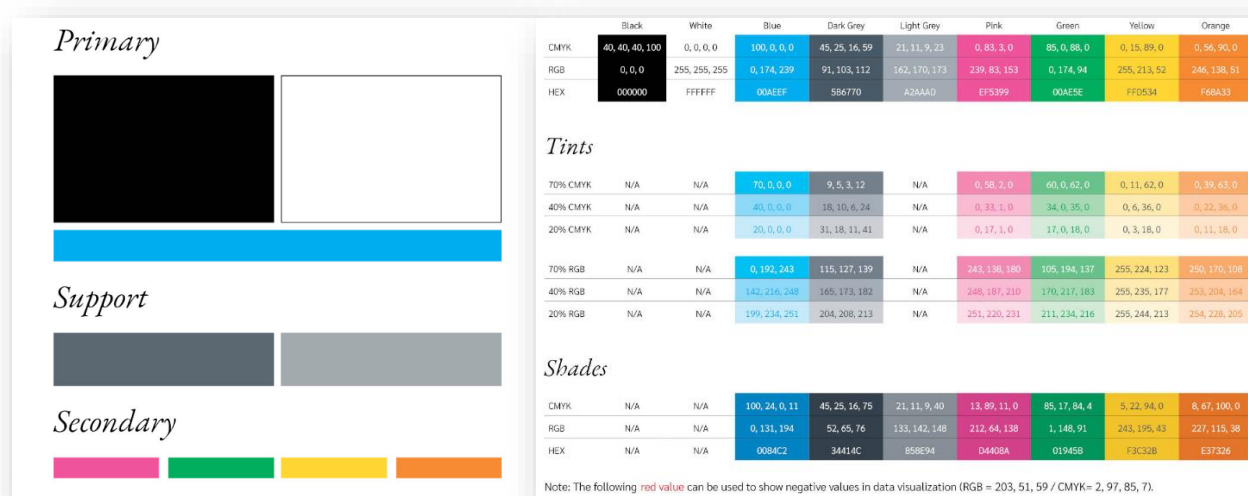
- "Aesthetics"
  - *aes()* in *ggplot*
  - Combine color with others aesthetics to ensure plot remains interpretable when printed black-and-white and for colorblind audience
  - There are also colorblind-proof palettes available, e.g., Okabe-Ito
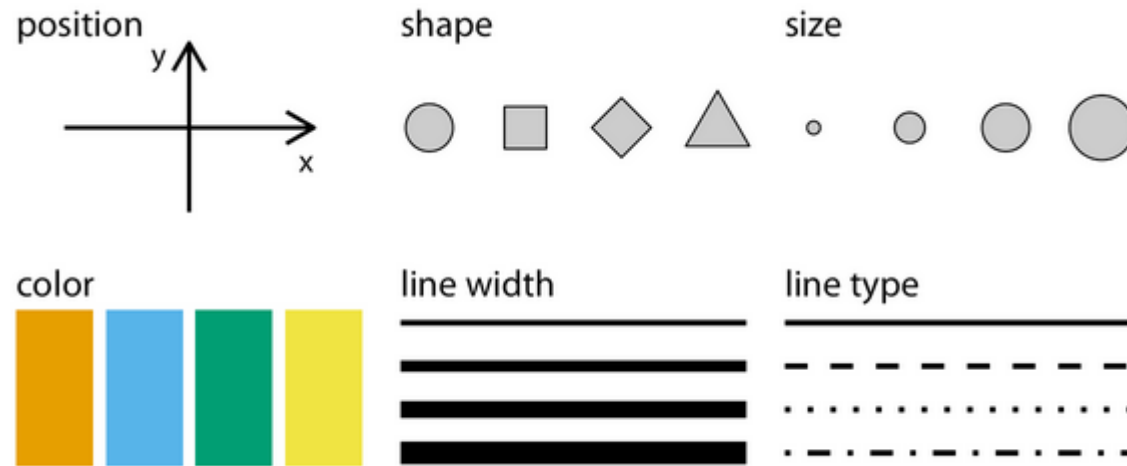
Color Considerations

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

- 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)
```
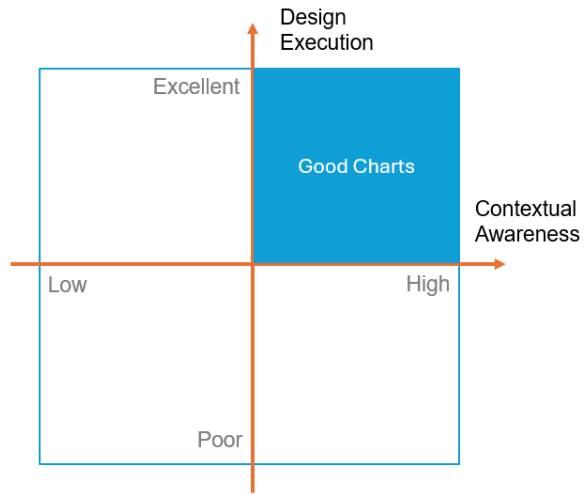
Visuaries design excellently and contextually

Design Execution

Contextual Awareness

Excellent

Good Charts

Low

High

Poor

Design
Execution

Excellent

Good Charts

Contextual
Awareness

Low   High

Poor
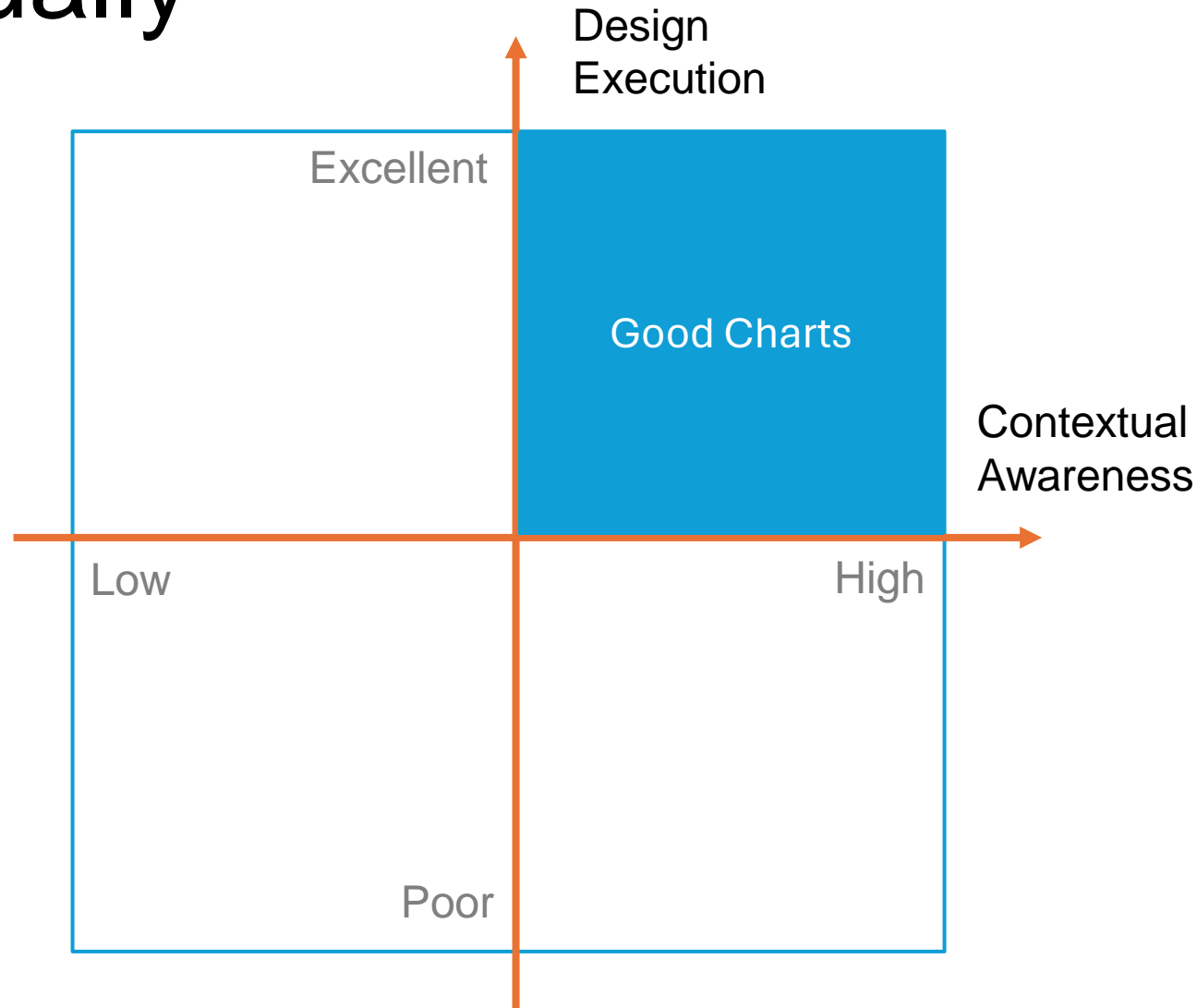
Design Execution

Reproducible R script available on GitHub

Original

Highway Miles Per Gallon vs. Engine Displacement
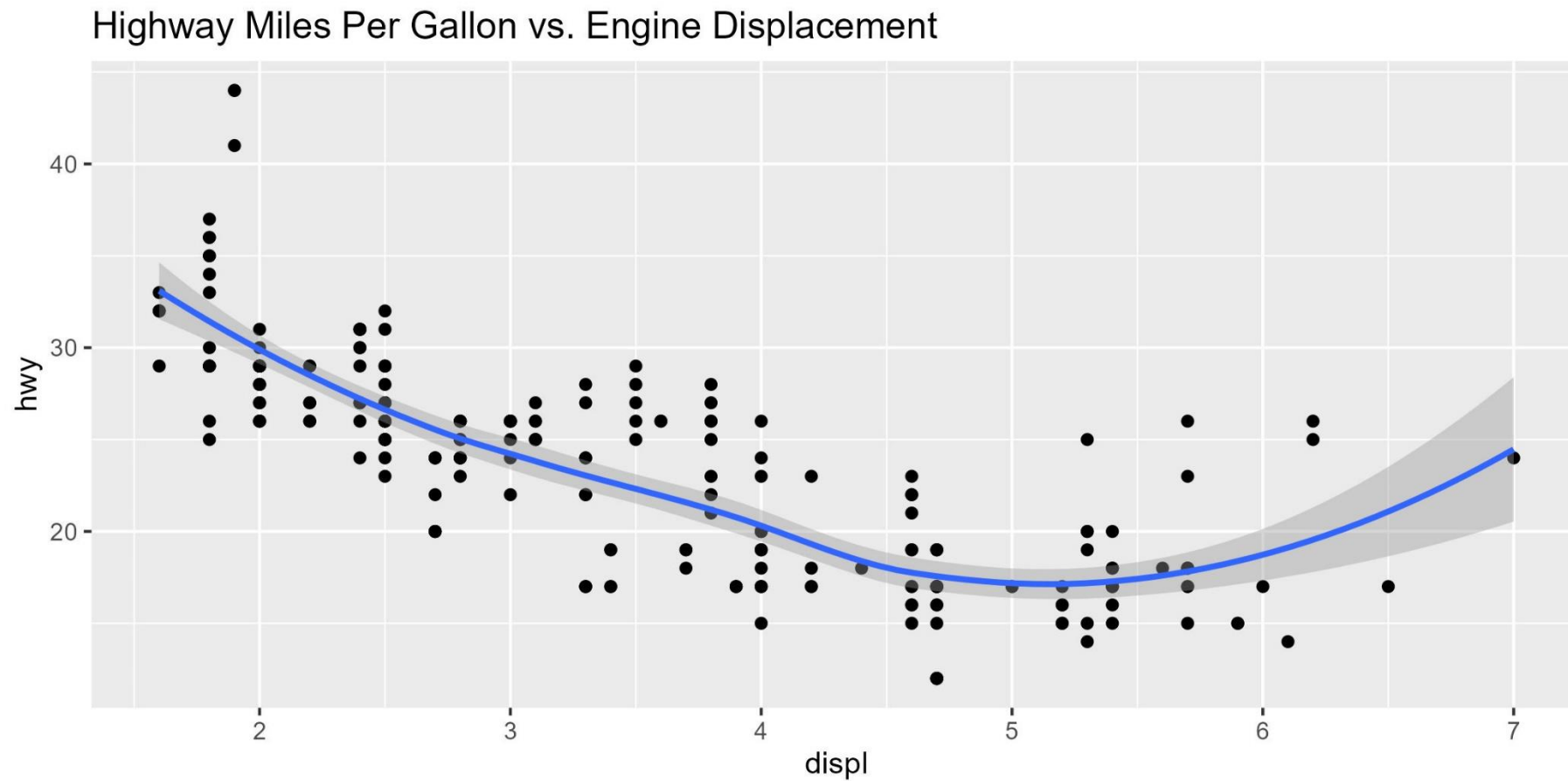
> How would you improve this plot?

Step 1: Look at your data!

```
> mpg
# A tibble: 234 × 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

Highway Miles Per Gallon vs. Engine Displacement
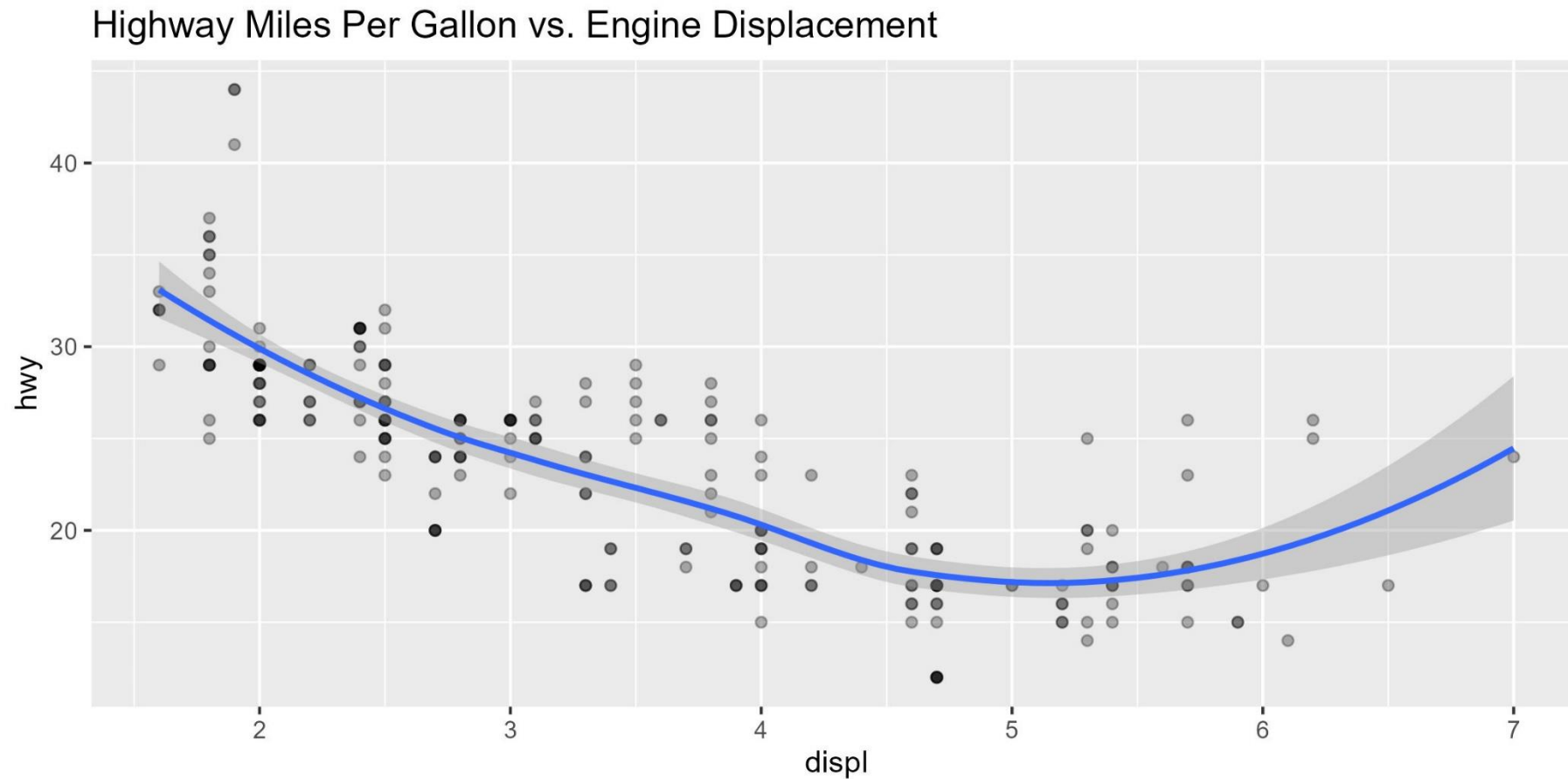
> "…above all else, show the data"

Intentional aesthetics - position

Highway Miles Per Gallon vs. Engine Displacement
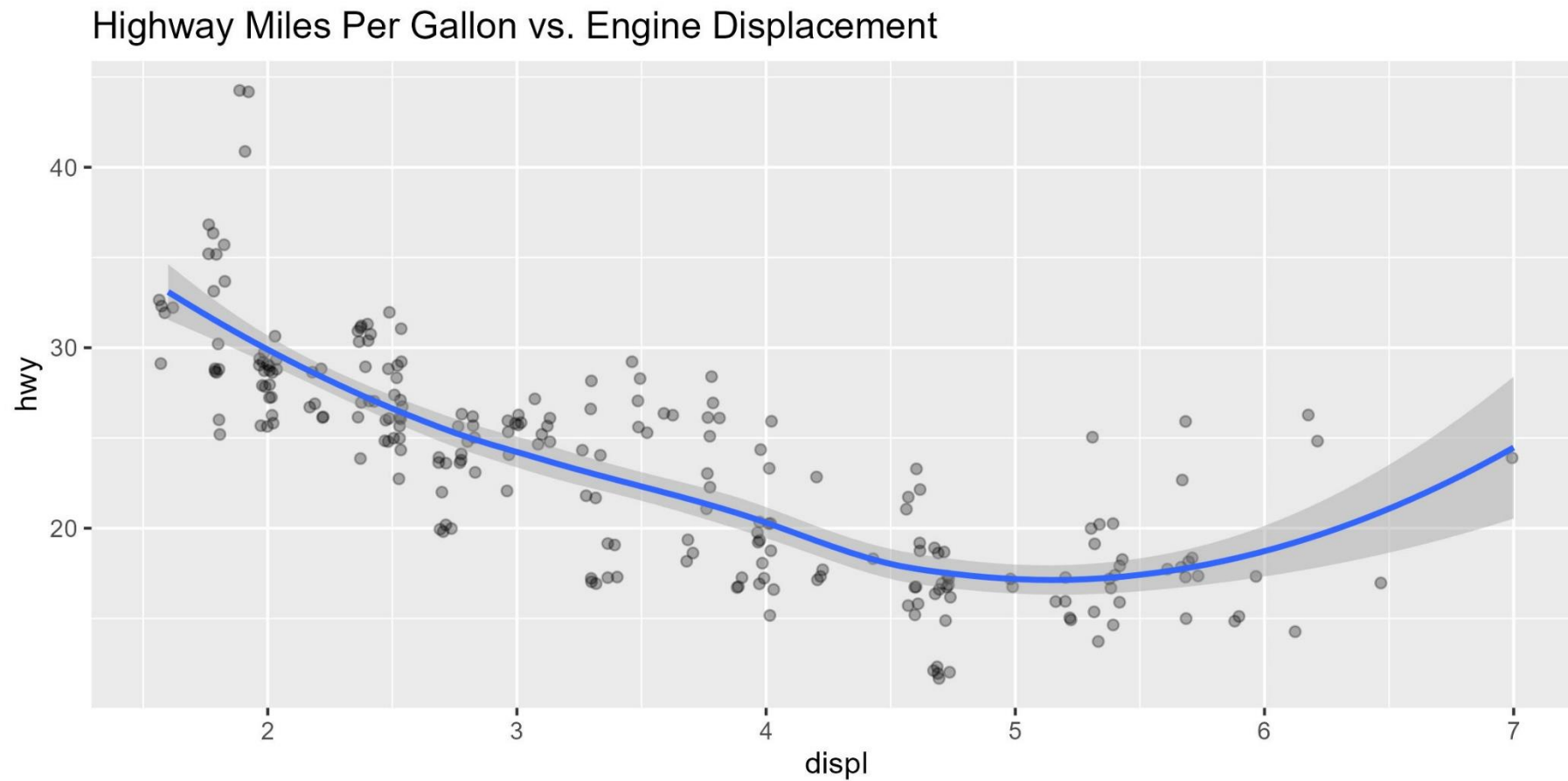
Descriptive labels and title

Fuel Efficiency Generally Decreases as Engine Displacement Increases

Remove distractions

Fuel Efficiency Generally Decreases as Engine Displacement Increases

Simpler trend line

Stick to the message

Fuel Efficiency Generally Decreases as Engine Displacement Increases
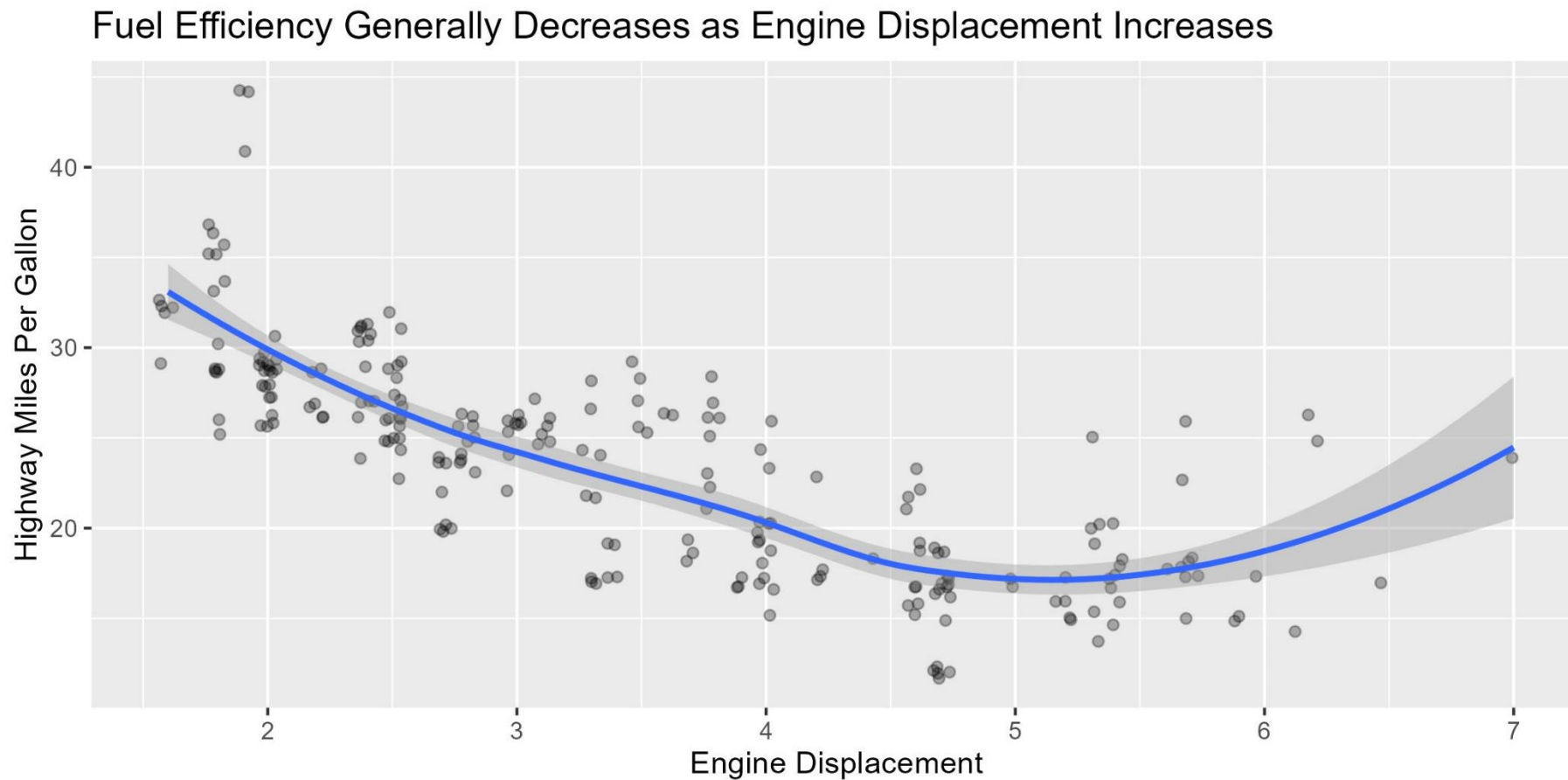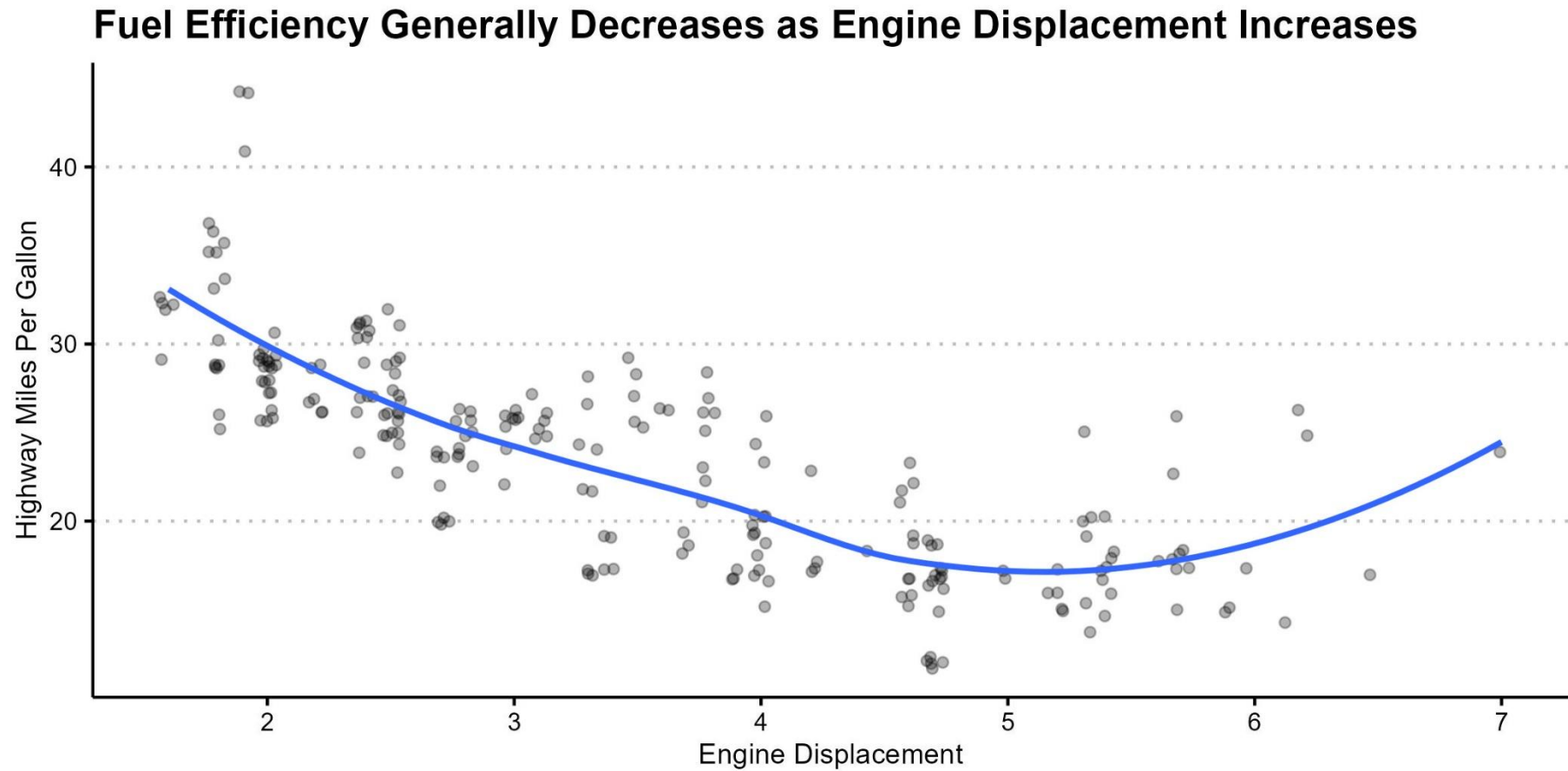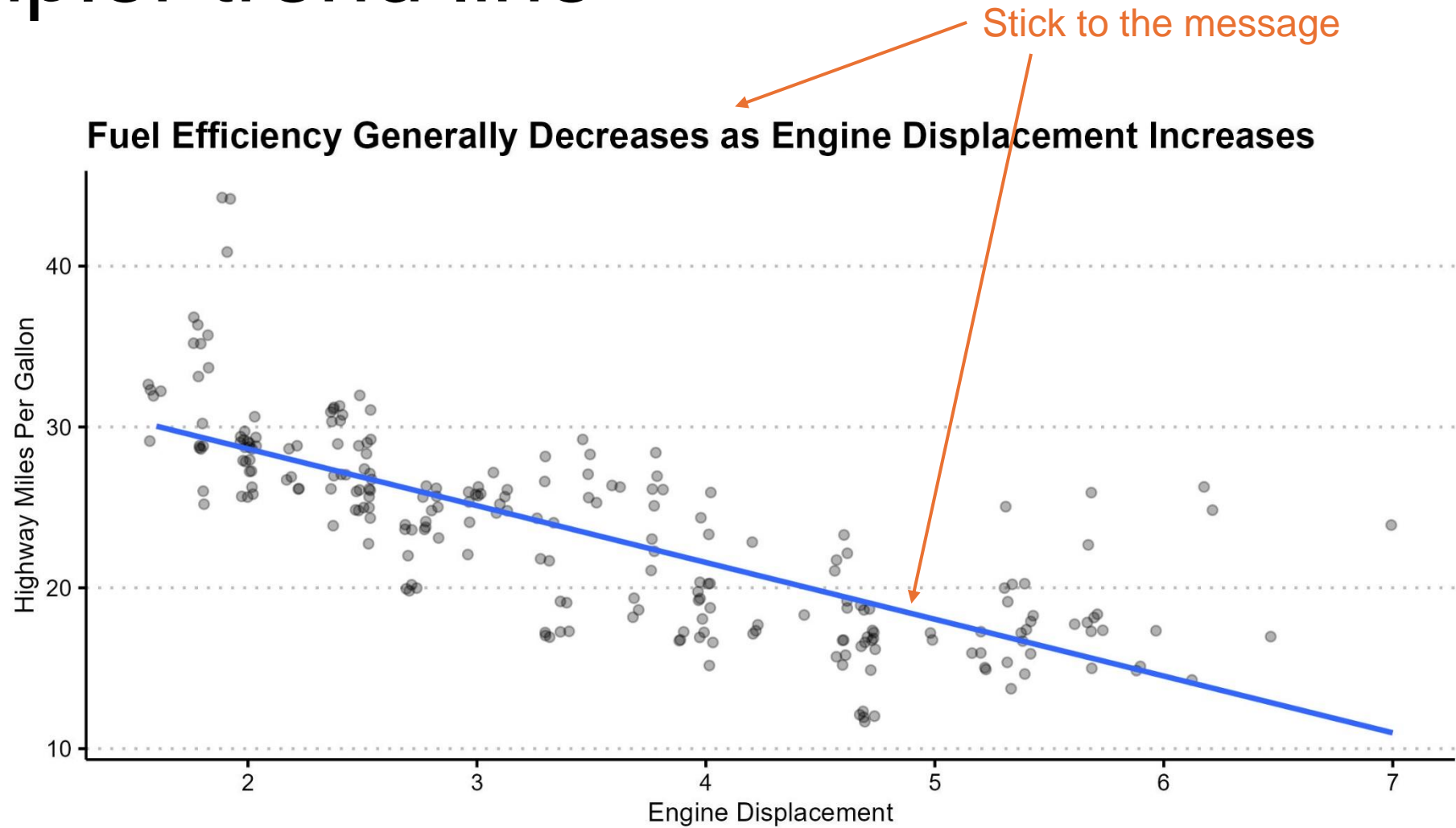
Increase data density – aesthetics

Fuel Efficiency Generally Decreases as Engine Displacement Increases
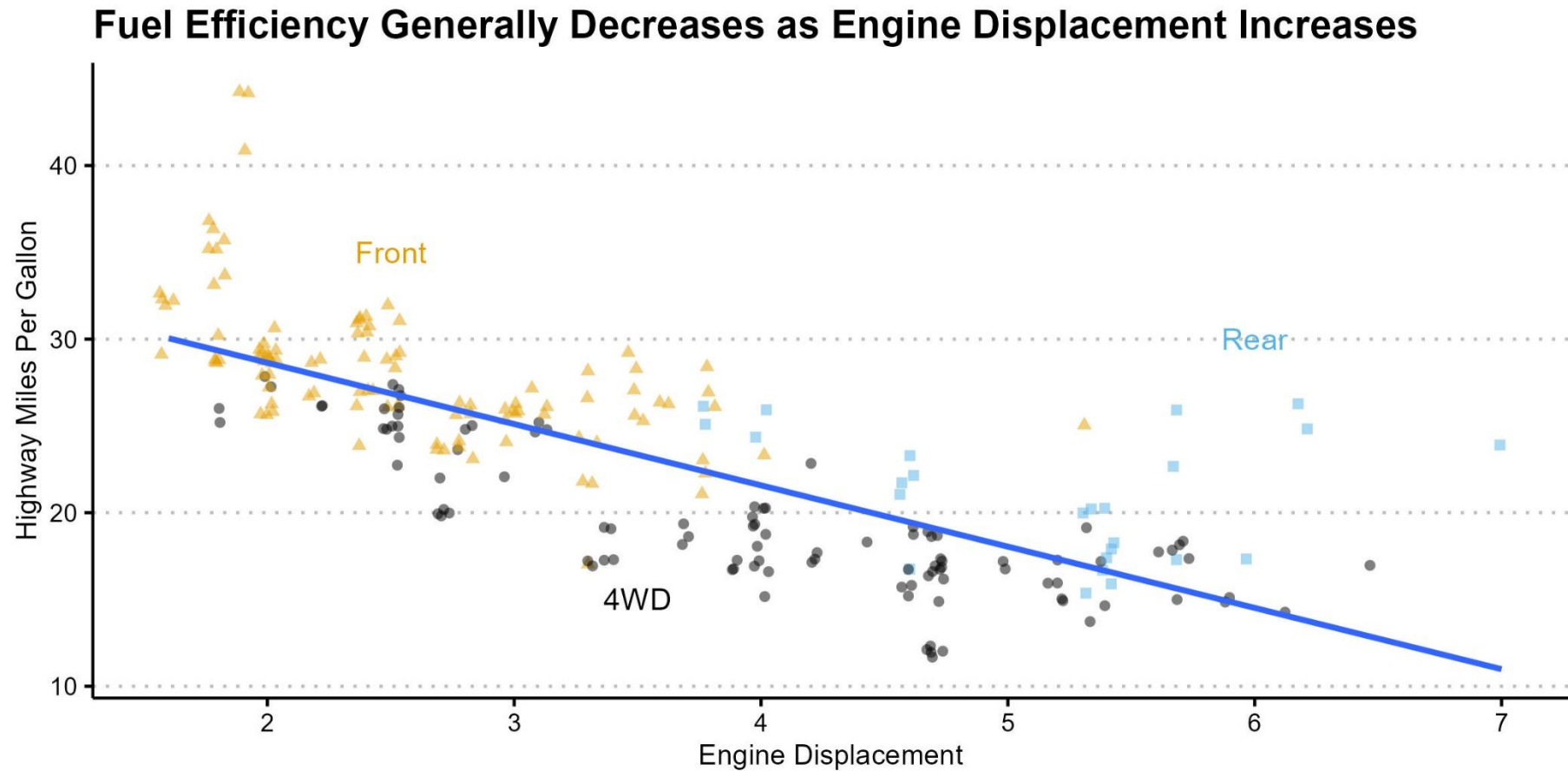
Increase data density – facets

Fuel Efficiency Generally Decreases as Engine Displacement Increases
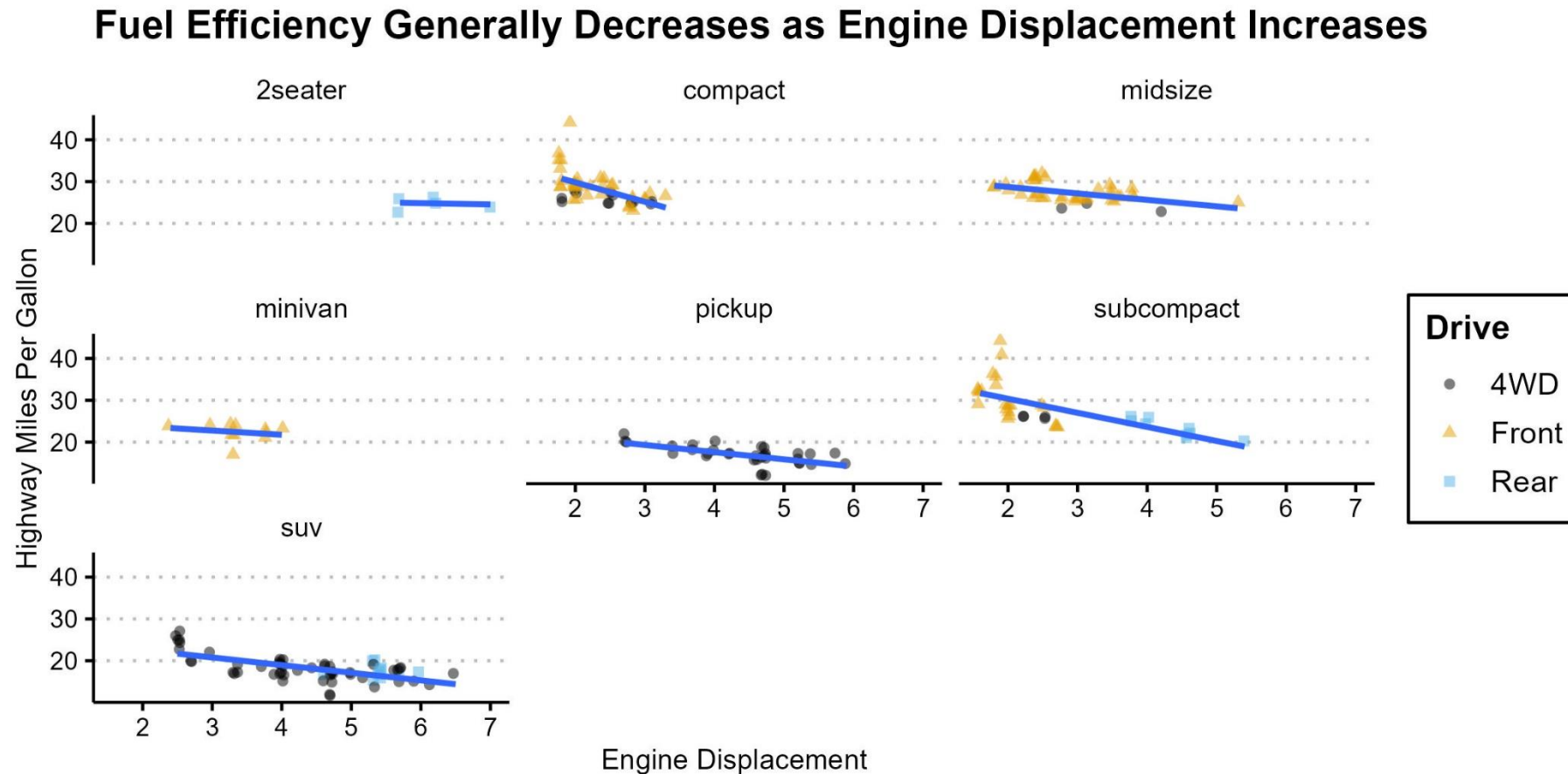
> Why should we not use alphabetical ordering for the panels?

Increase data density – order matters

Fuel Efficiency Generally Decreases as Engine Displacement Increases
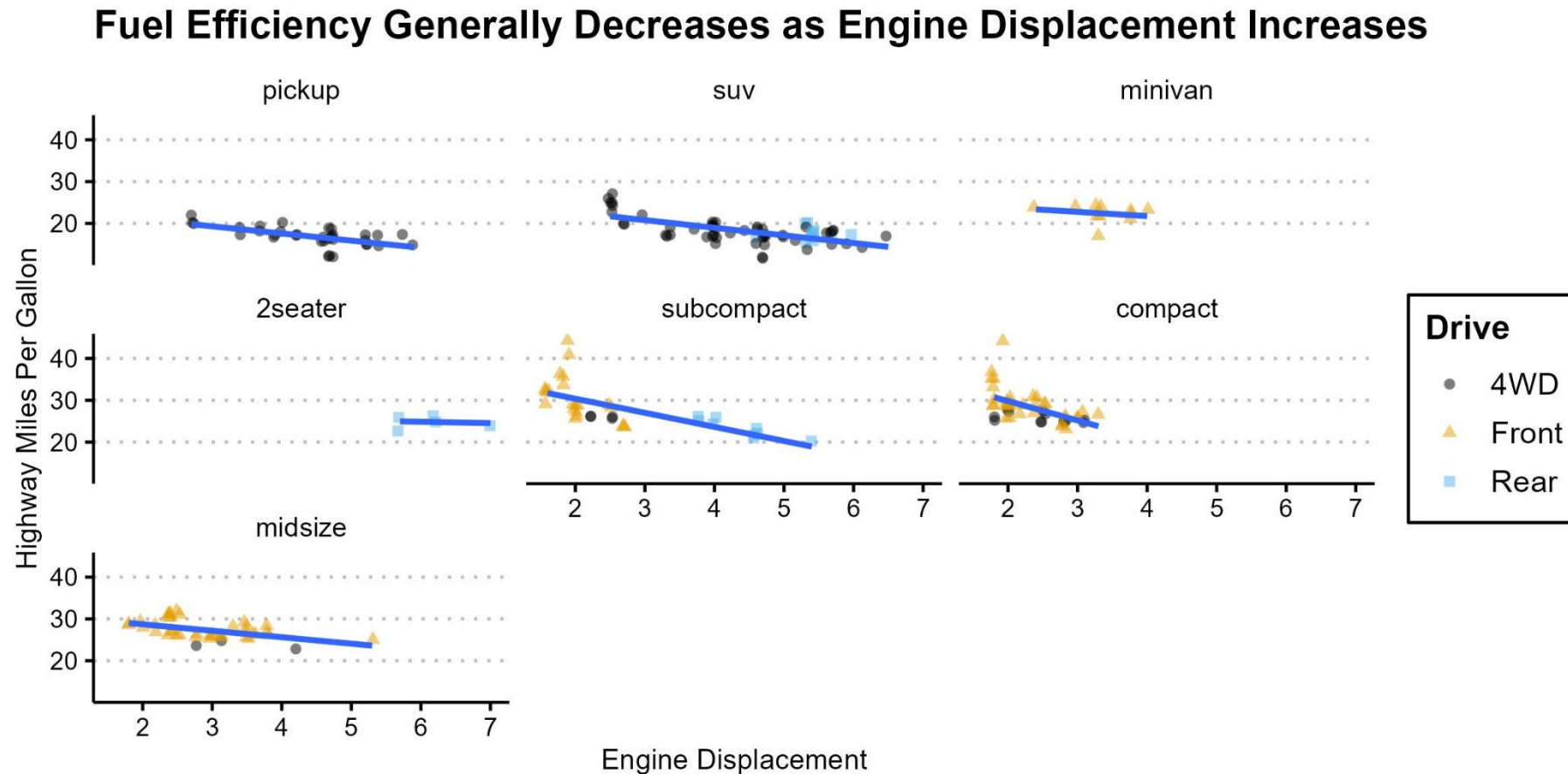
> Order panels by median Highway MPG of each class

Labels - static

Fuel Efficiency Generally Decreases as Engine Displacement Increases
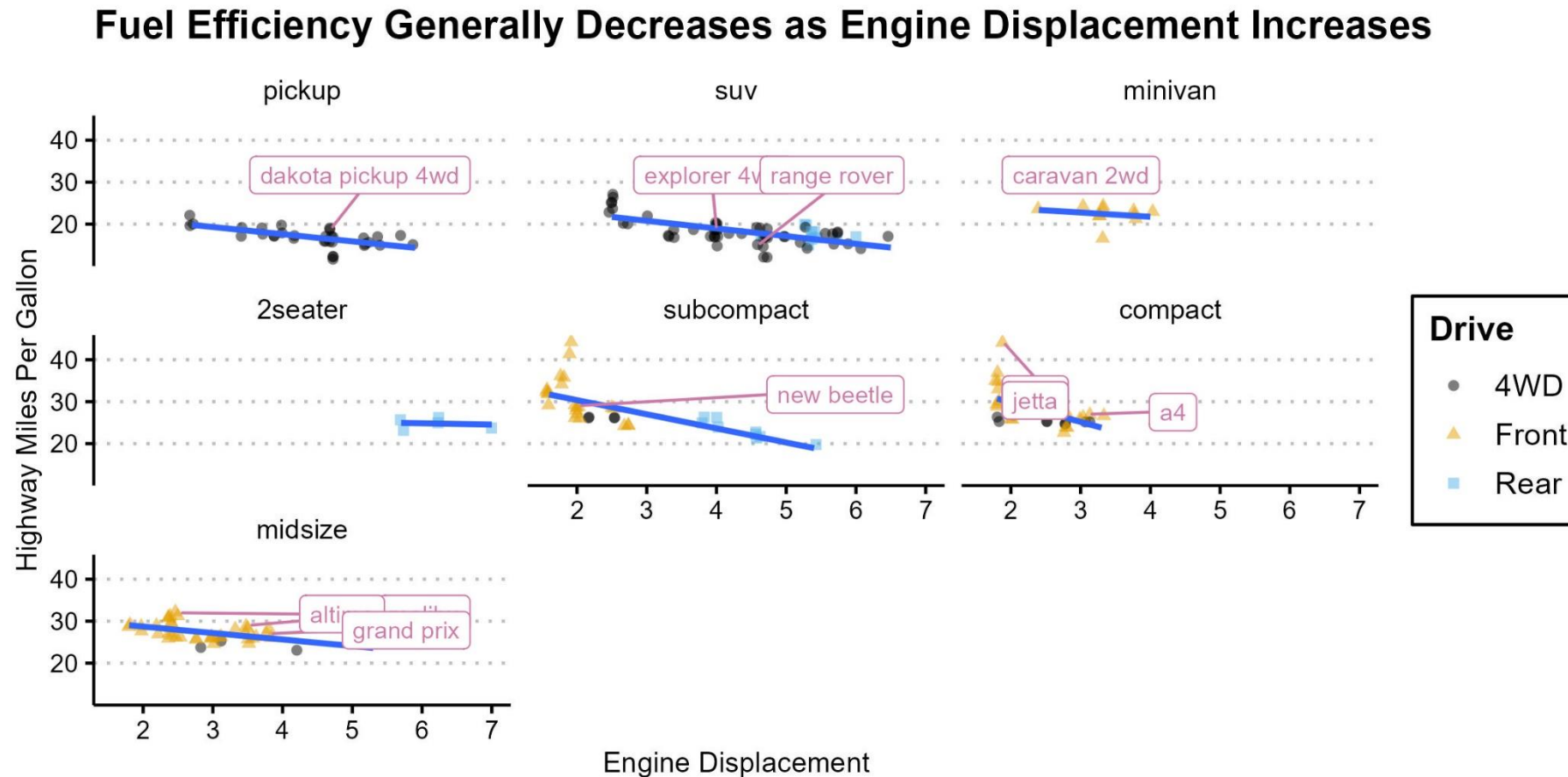
> Helpful but too busy. Anything else we could do?

Labels - interactive

Fuel Efficiency Generally Decreases as Engine Displacement Increases
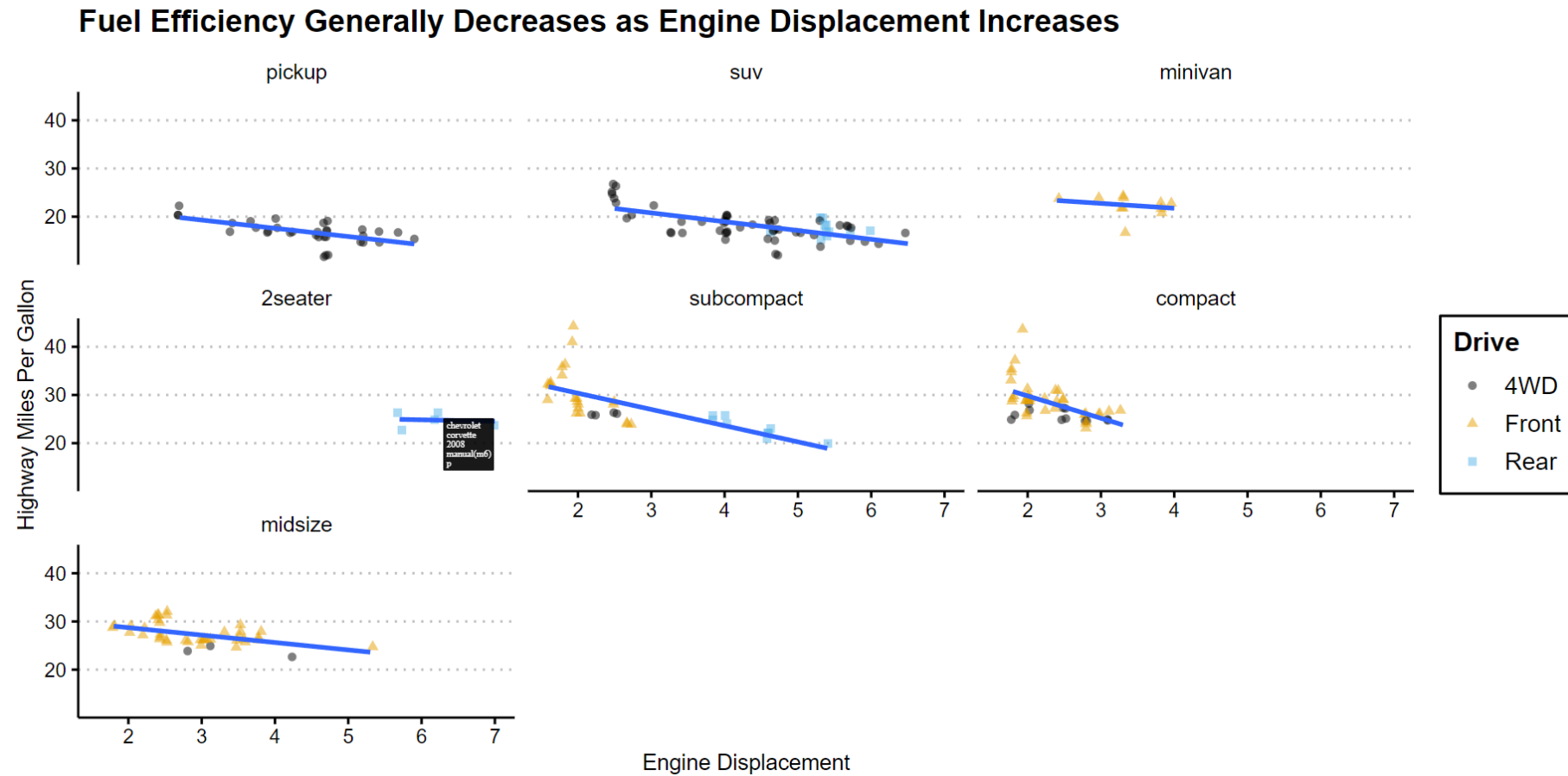
Poll: Which of the following charts appeals to you the most?
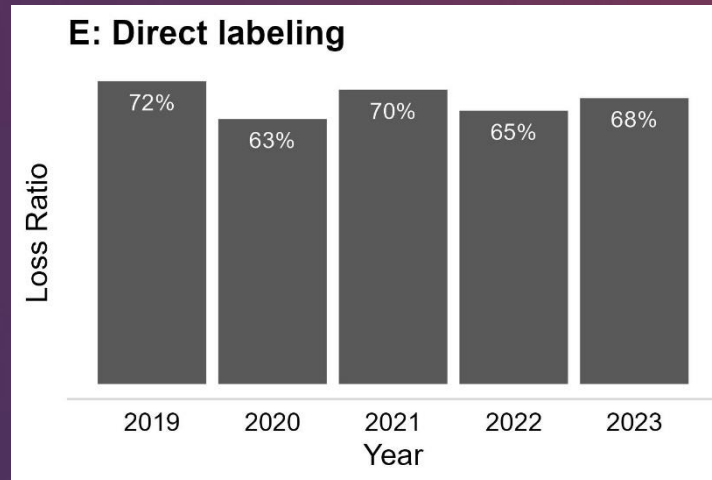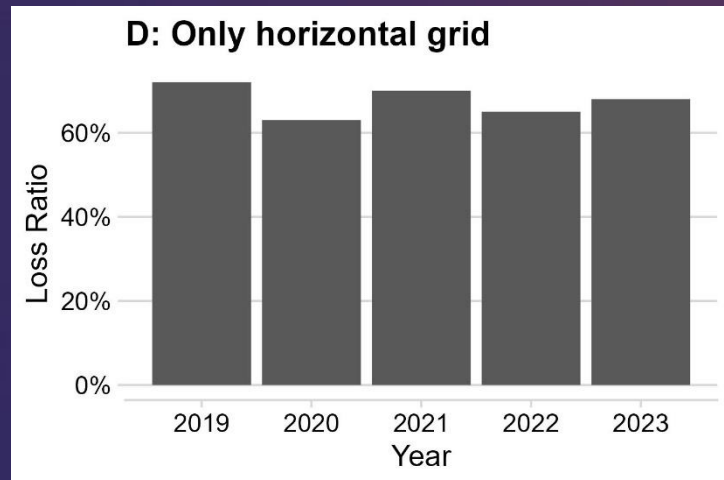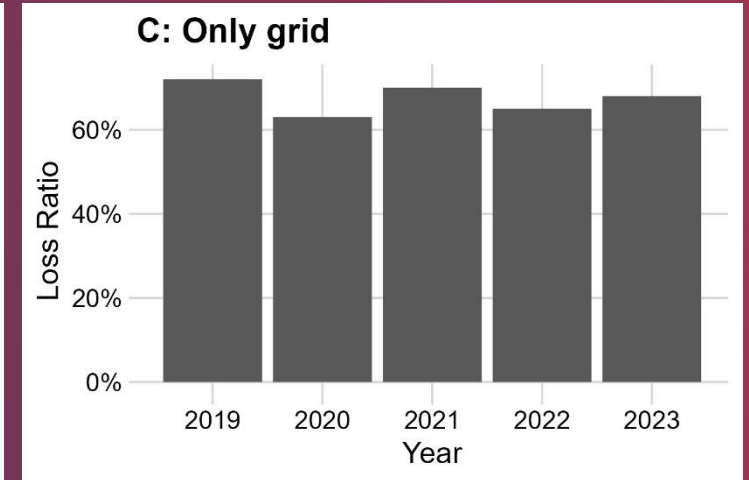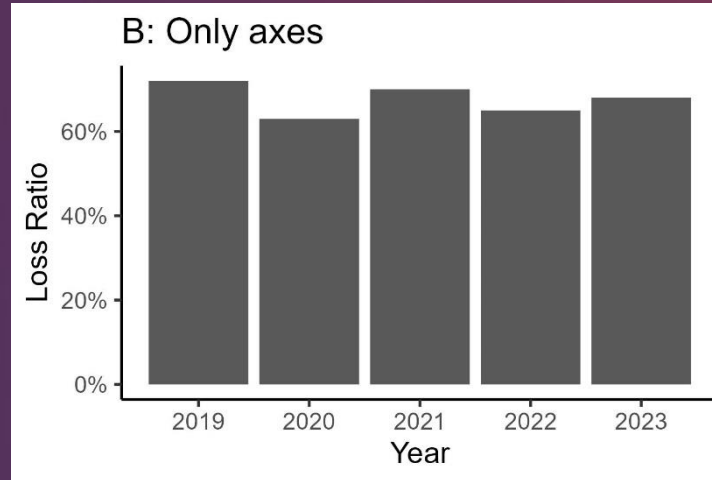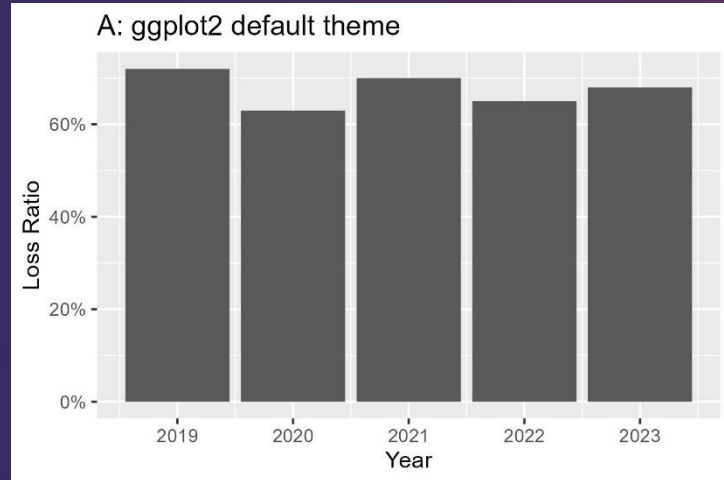
This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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?
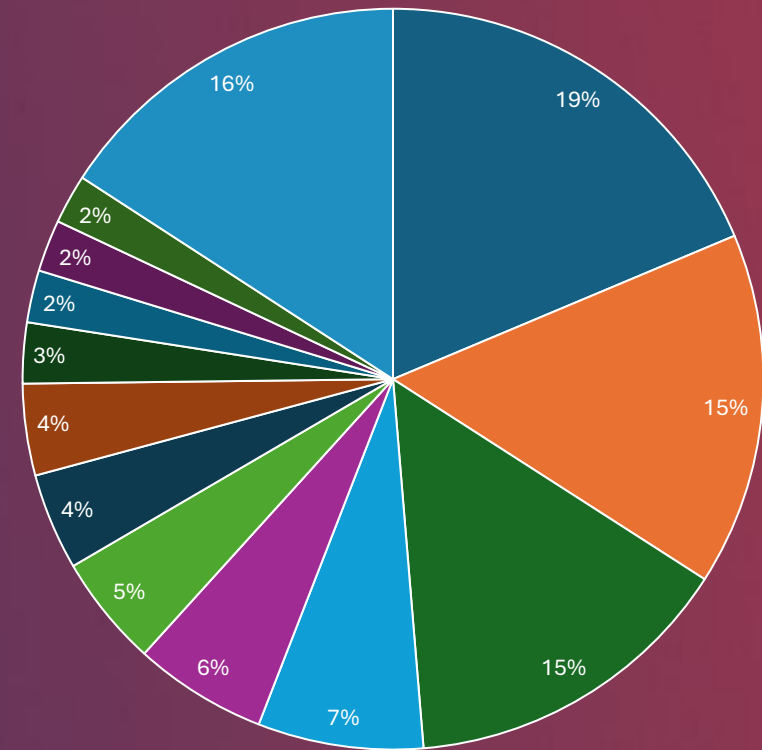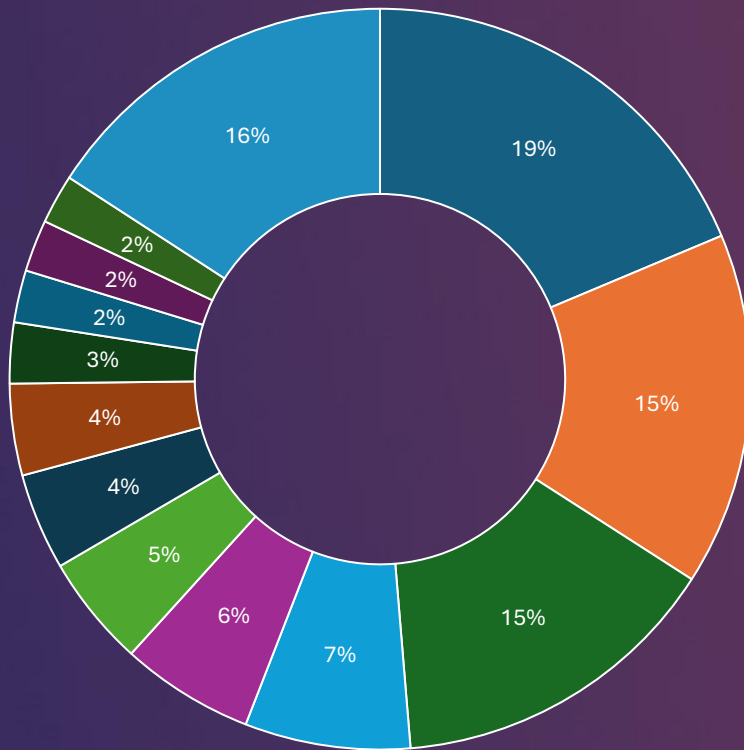
Poll: Should doughnuts and pies be avoided?

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

This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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"

• Amounts

The "what"

- Single distribution

Histogram

Density Plot

Cumulative Density

A CUMULATIVE FREQUENCY CURVE
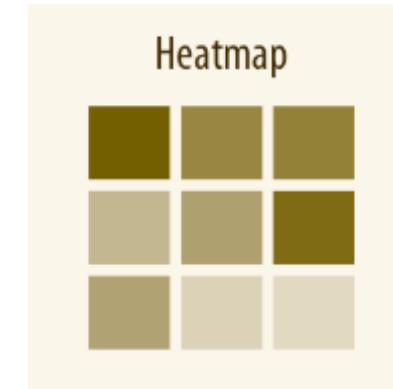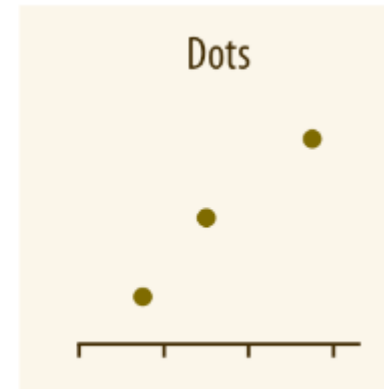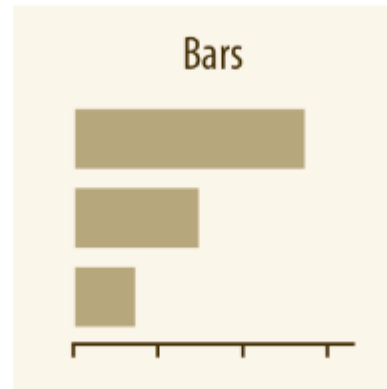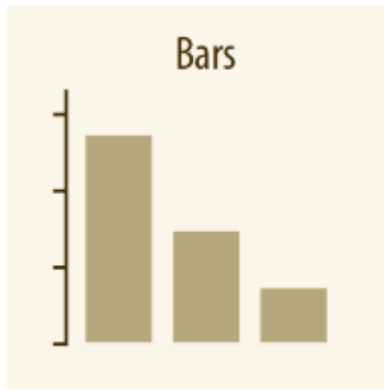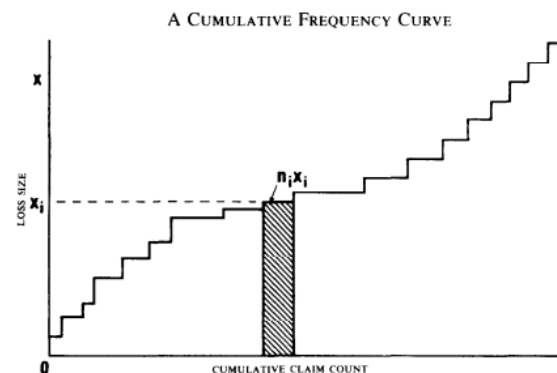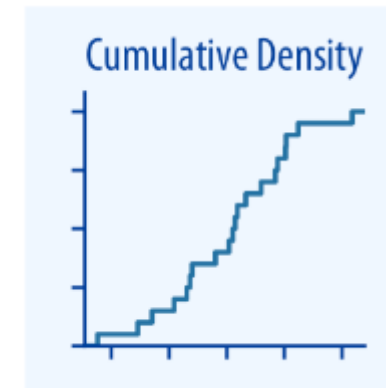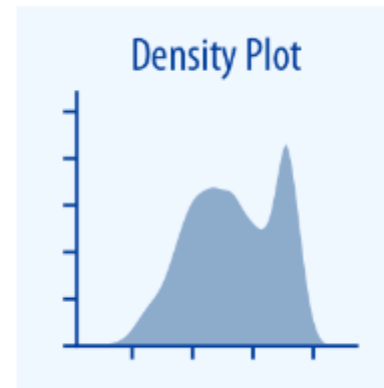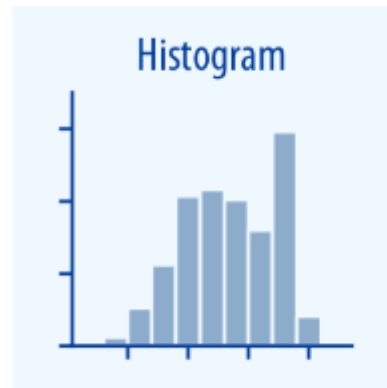
Fundamentals of Data Visualization (clauswilke.com)

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

The "what"

- Multiple distributions
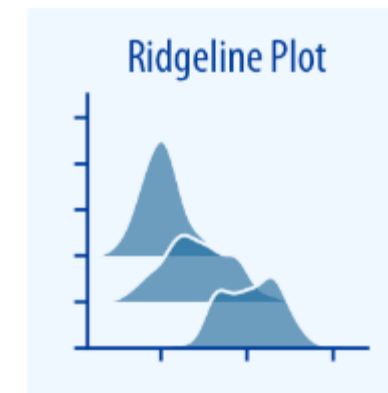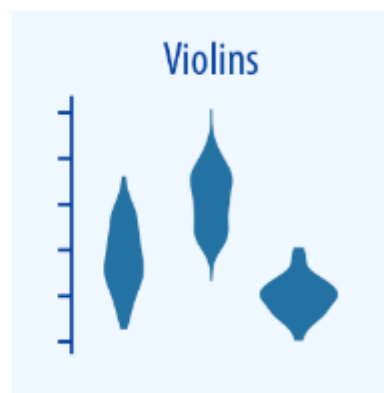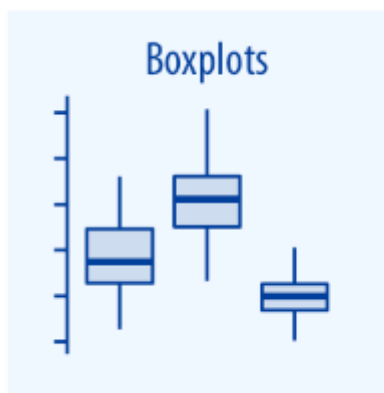
The "what"

- Proportions

Pie Chart

Bars

Stacked Bars

Treemap

The "what"

- Multivariate relationships

Scatterplot

Slopegraph

2D Bins

Correlogram

Fundamentals of Data Visualization (clauswilke.com)

The "what"

- Geospatial

Map

Choropleth

Cartogram

Cartogram Heatmap
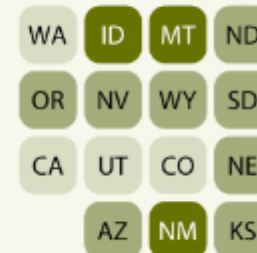
WA	ID	MT	ND
OR	NV	WY	SD
CA	UT	CO	NE
	AZ	NM	KS

Fundamentals of Data Visualization (clauswilke.com)

The "what"

• Uncertainty
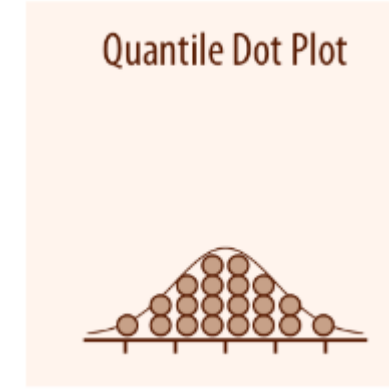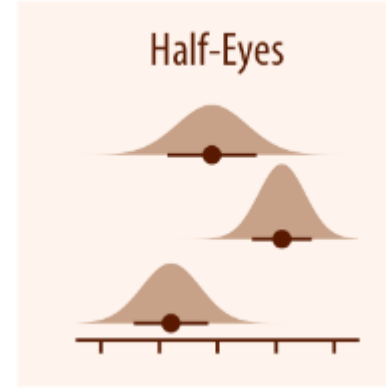
Poll: You want to communicate that a distribution is more heavy-tailed than the other. Which plot below would you choose for an actuarial audience?
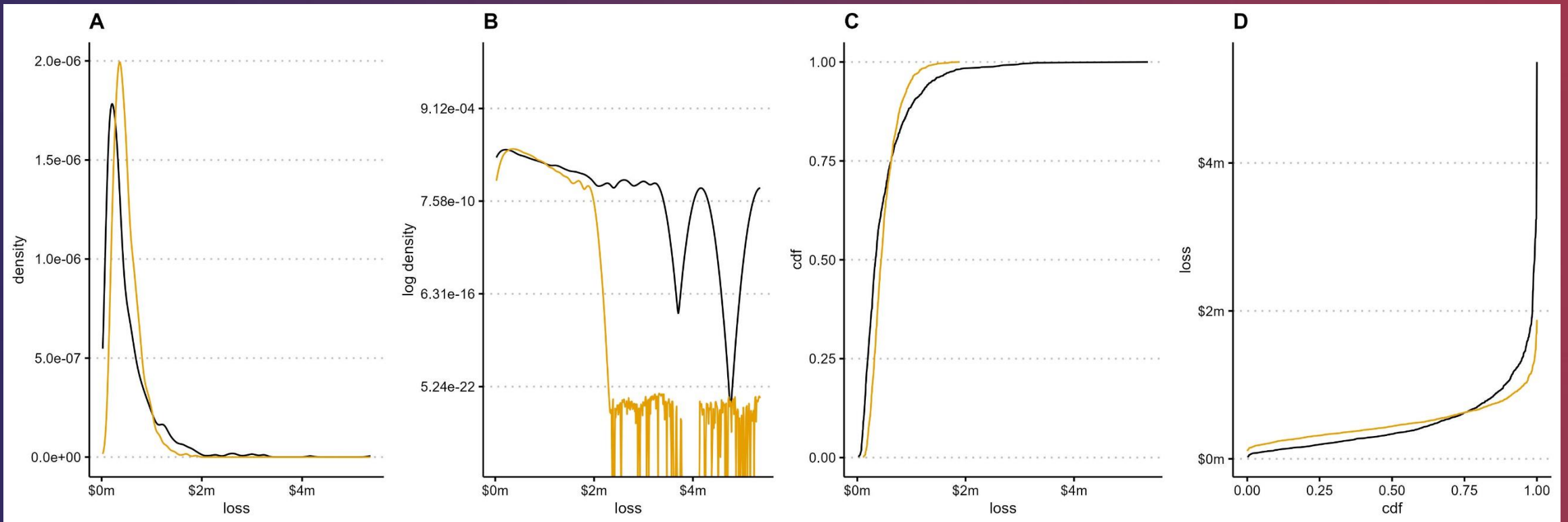
This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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?
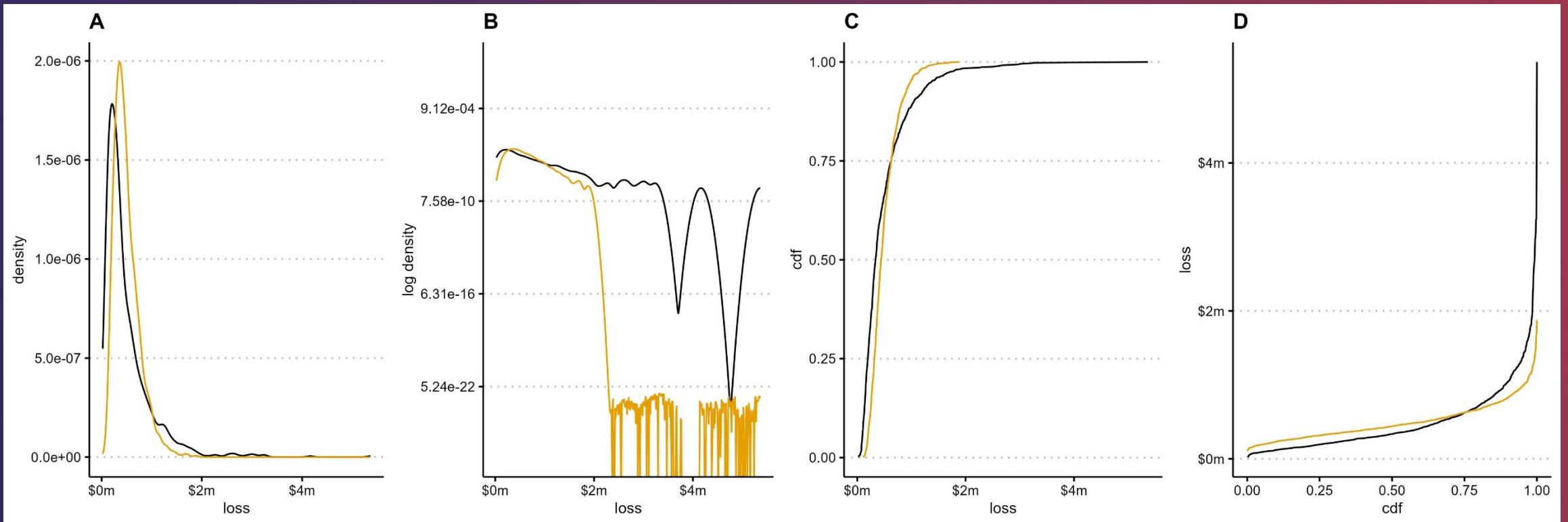
This slide will activate the polling question once the presentation is started.
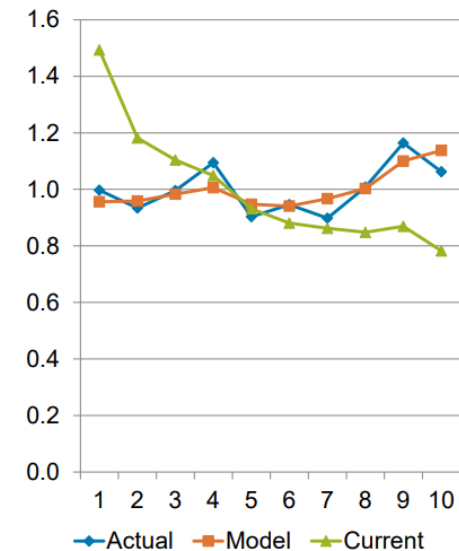Do not change or delete text on this slide.

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?
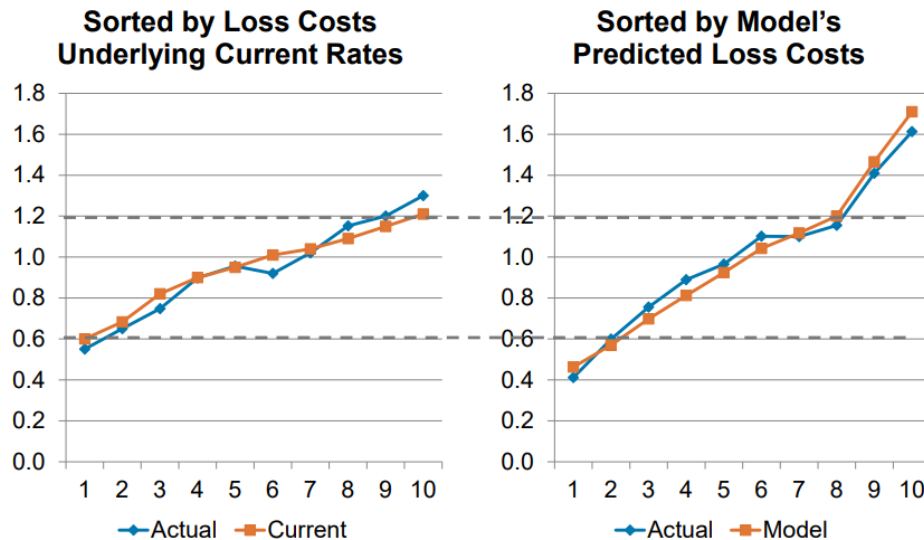
This slide will activate the polling question once the presentation is started.
Do not change or delete text on this slide.

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

Thank you!

✈ enbo.jiang@lockton.com

in 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

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