

# Proposed Content

**Title:** Effective Data Visualization for Actuaries

**Session Description:**

As actuaries, we are expected to understand the structure and rationale behind complex statistical models. Beyond that, we need to evaluate whether one model is preferred over another in a well-defined business context. Finally, these conclusions must be shared with stakeholders who are then expected to digest and support the actuary's conclusions. Although data visualization is often touted as having maximal benefit for the last step of the process, we contend that it provides value throughout

In this session, Brian Fannin and Jordan Bonner will highlight key actuarial concepts and exhibits that can be better understood through data visualization. R's `ggplot2` package will be used to create exhibits that allow for a deeper level of understanding and intuition. By growing comfortable with the value and mechanics of visualization at all stages of the analytics cycle, actuaries will become more facile in communicating with any stakeholder at any stage.

**Learning Objectives:**

1. Unlock a deeper understanding of complicated concepts through data visualization
2. Leverage R's `ggplot2` package to create compelling actuarial exhibits
3. Restructure data visualizations to tell the right story for your audience

**To Be Removed. For Reference Only.**



# Ideas for Interactivity

- Familiarity with R
- “Which chart is easier to interpret”
- Test Bar/Line bias (<https://bit.ly/3ymQ1eE>)



# Effective Data Visualization for Actuaries

Brian Fannin

Jordan Bonner



# 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.



# Agenda

- **Seeing is Believing**
- **Decoding Complexity**
- **The Power of Persuasion**





**SEEING IS BELIEVING**



# Seeing is Believing

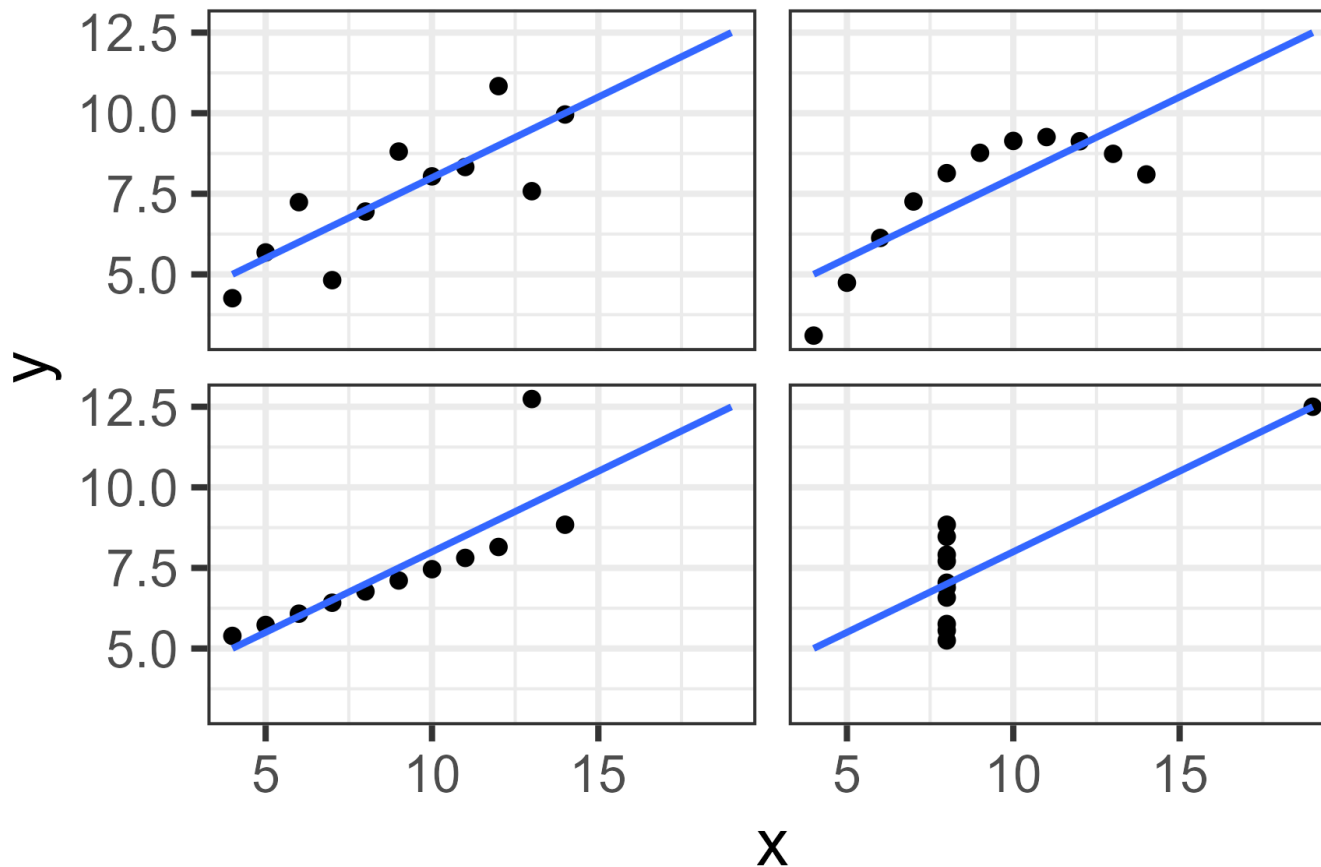
**Imagine you have two sets of points with similar:**

- means ( $\bar{x}$ ,  $\bar{y}$ )
- sample variances ( $s_x^2$ ,  $s_y^2$ )
- correlation ( $\rho_{xy}$ ),
- coefficient of determination ( $R^2$ )
- regression line

**How similar might these datasets be?**



# Anscombe's Quartet



**Descriptive  
Statistics  
Shared:**

$$\bar{x}, \bar{y}$$

$$s_x^2, s_y^2$$

$$\rho_{xy}$$

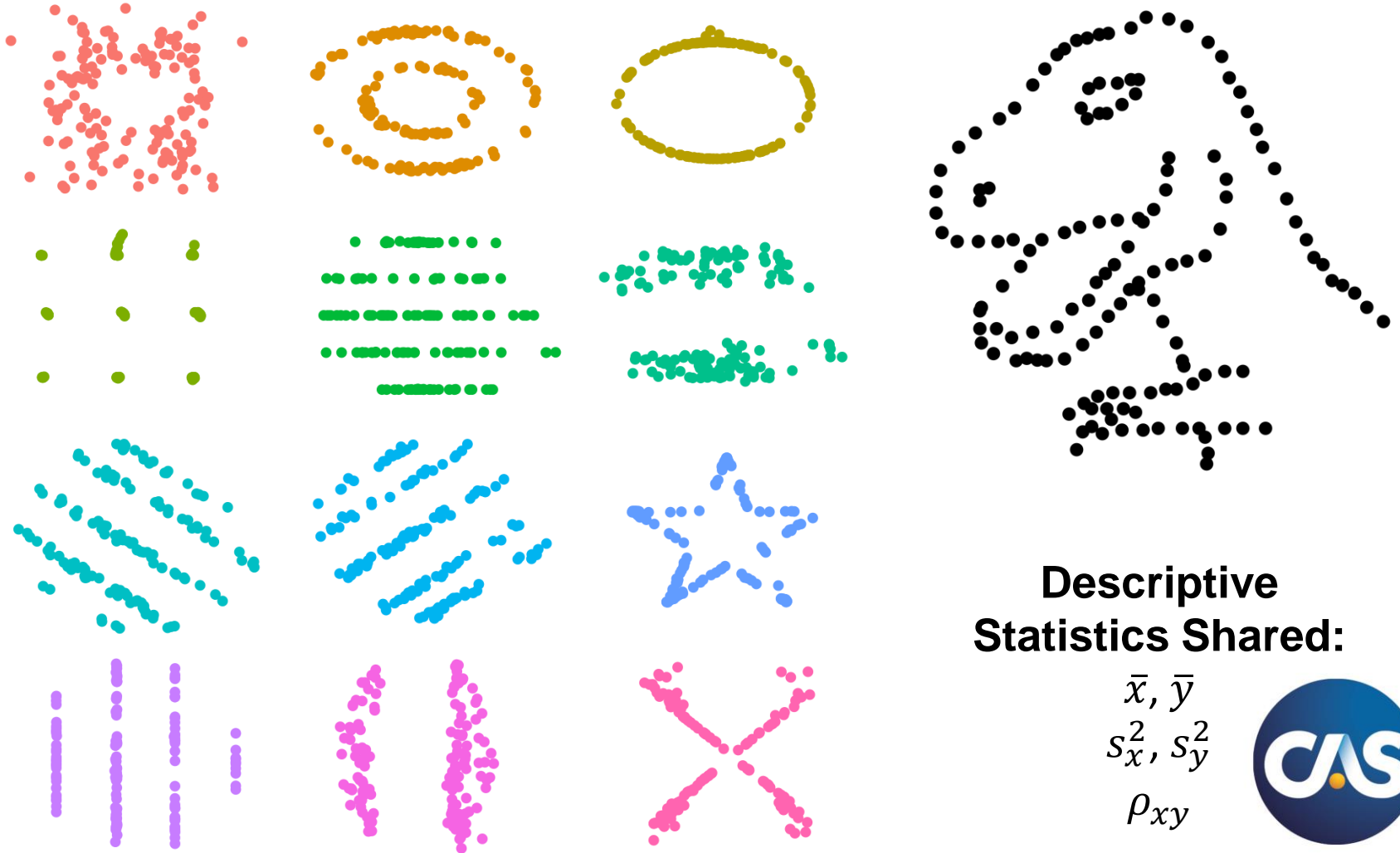
$$R^2$$

$$\hat{\beta}_0 \text{ and } \hat{\beta}_1$$





# The Datasaurus Dozen



# Why Seeing is Believing:

- **Better Pattern Identification**
- **“Big Picture”**
- **Identifying Outliers**
- **Easily Interpreted**
- **Memory Retention**



# DECODING COMPLEXITY



# Decoding Complexity

- Univariate Linear Regression
- Principal Components Analysis
- Decision Trees
- Correlation & Tail Correlation



# THE POWER OF PERSUASION



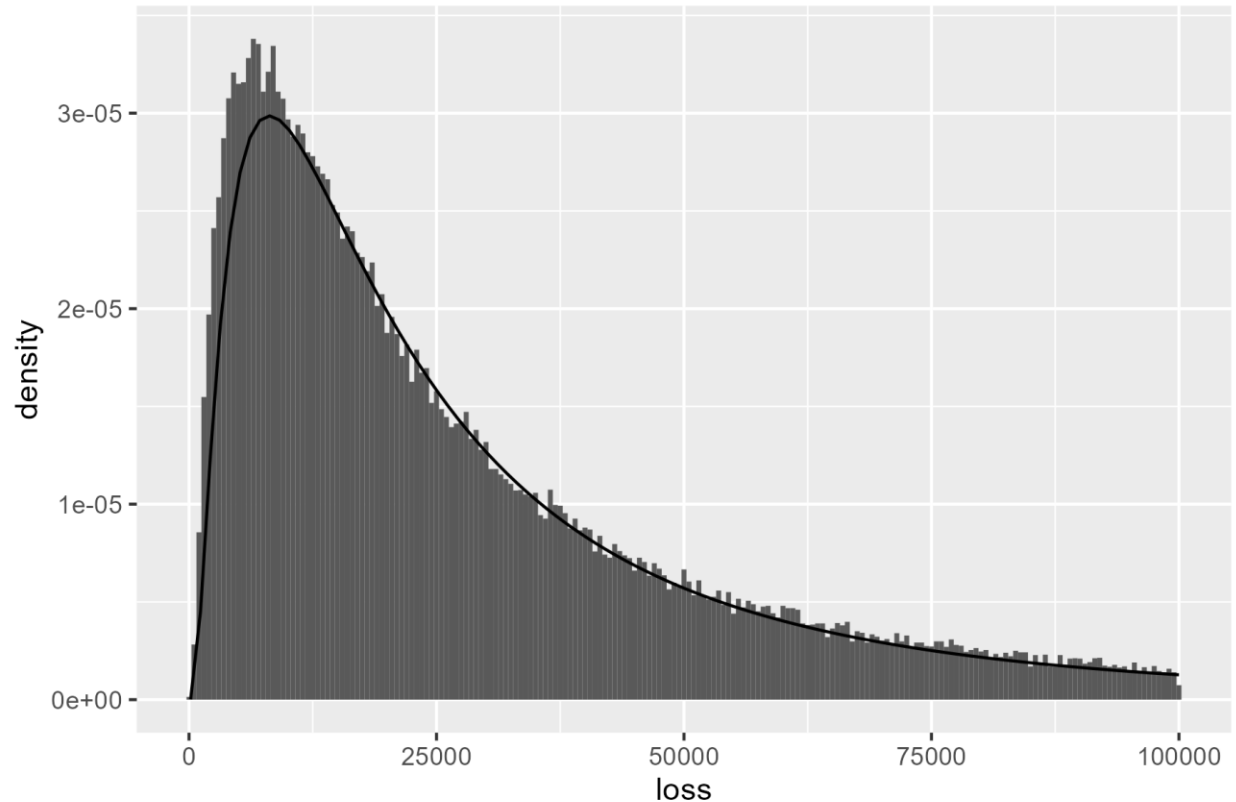
# The Power of Persuasion

- Highlighting Key Insights
- Telling a Story
- Providing Context
- Show Comparisons



# Initial Example: Fitted Severity

How can we  
improve upon  
this chart?

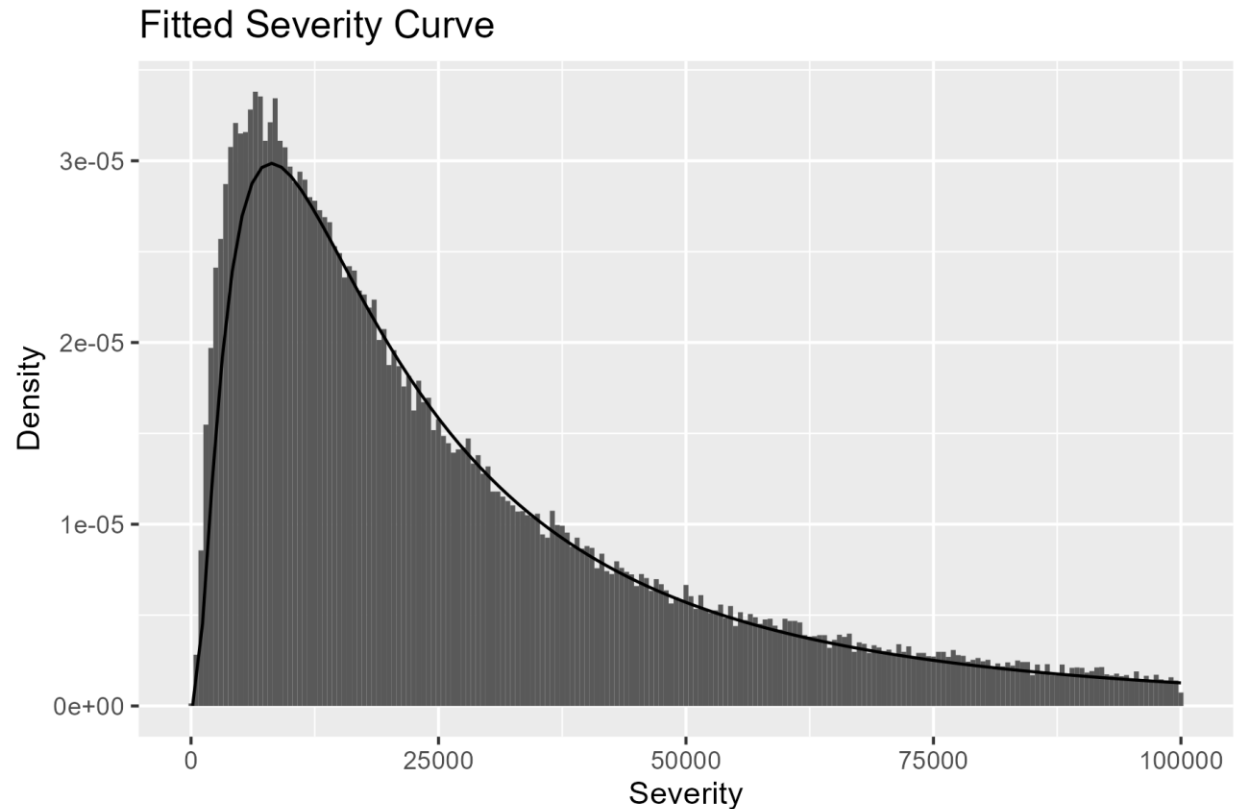


```
p <- data |>  
  geom_histogram(aes(x = loss, after_stat(density))) +  
  stat_function(fun = dlnorm,  
               args = list(meanlog = 10, sdlog = 1.0))
```



# Titles & Renaming Variables

Adding a title  
and variable  
names goes a  
long way.



```
p <- p +  
  labs(  
    x = "Severity",  
    y = "Density",  
    title = "Fitted Severity Curve"  
  )
```

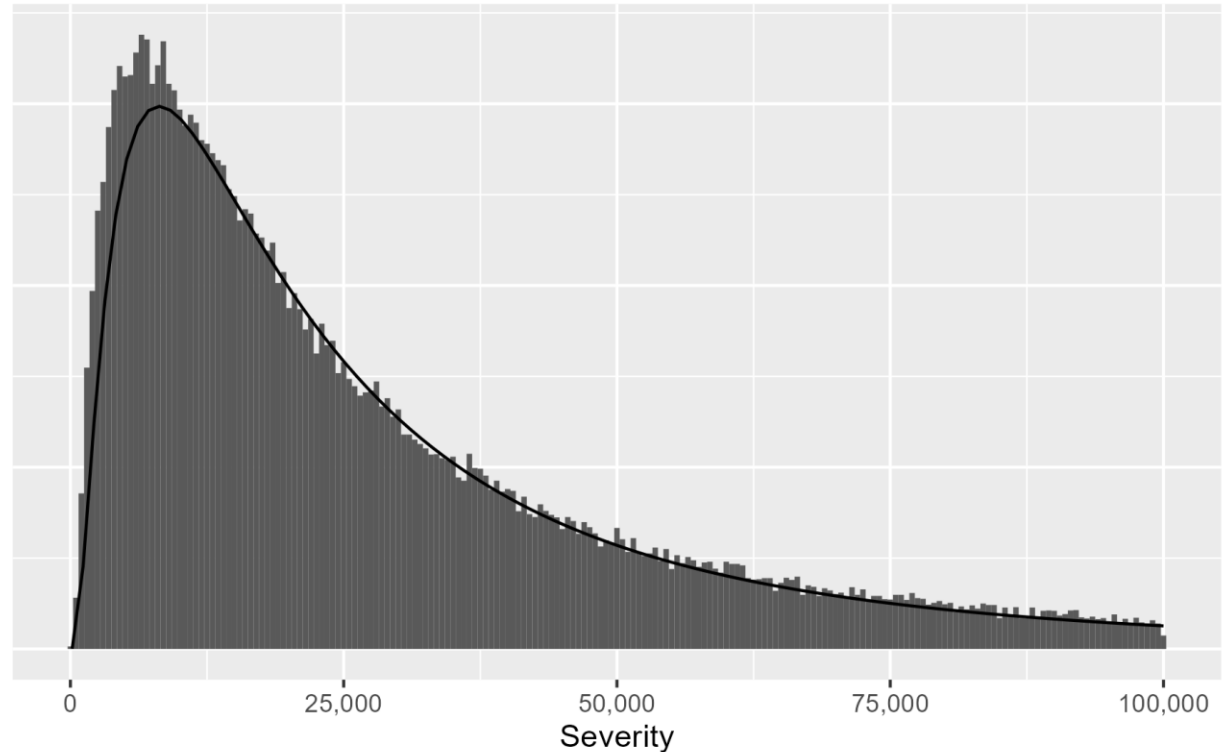




# Cleaning Axes

We can  
change the x-  
axis labels  
and remove  
the y-axis.

Fitted Severity Curve



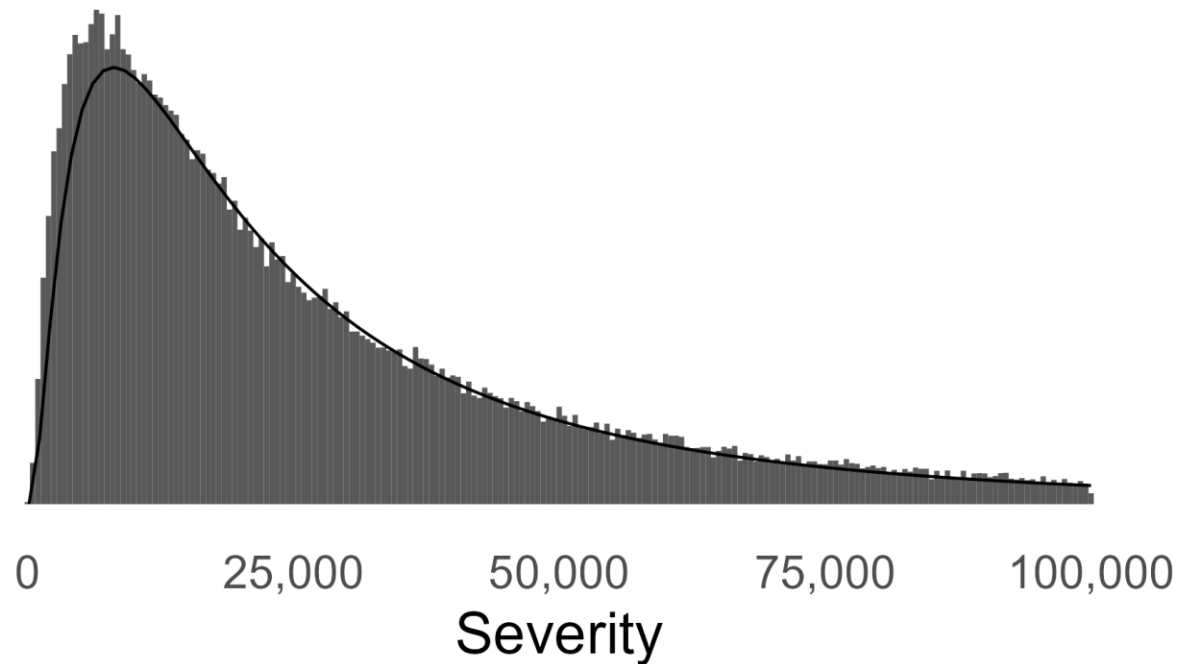
```
p <- p +  
  scale_x_continuous(labels = scales::comma) +  
  theme(axis.text.y = element_blank(),  
        axis.ticks.y = element_blank(),  
        axis.title.y = element_blank())
```



# Text Size & Theme

Fitted Severity Curve

We can increase text size and select a simpler theme.



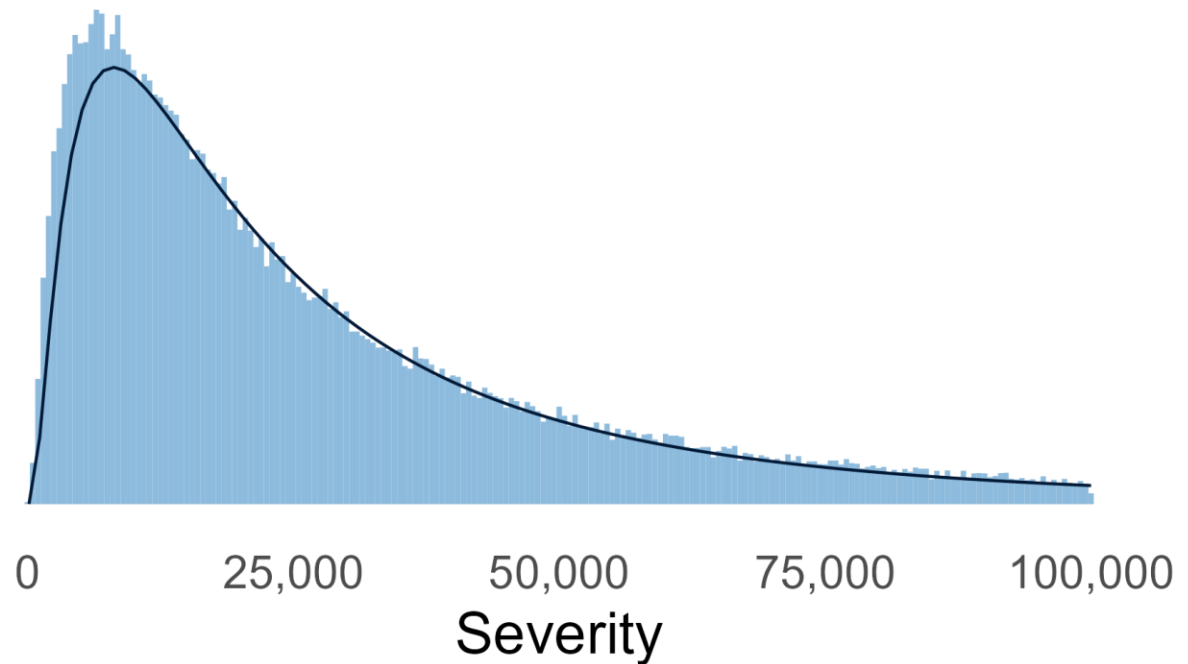
```
p <- p +  
  theme_minimal(base_size = 20) +  
  theme(panel.grid.major = element_blank(),  
        panel.grid.minor = element_blank()) +  
  scale_x_continuous(labels = scales::comma) +  
  theme(axis.text.y = element_blank(),  
        axis.ticks.y = element_blank(),  
        axis.title.y = element_blank())
```



# Color

## Fitted Severity Curve

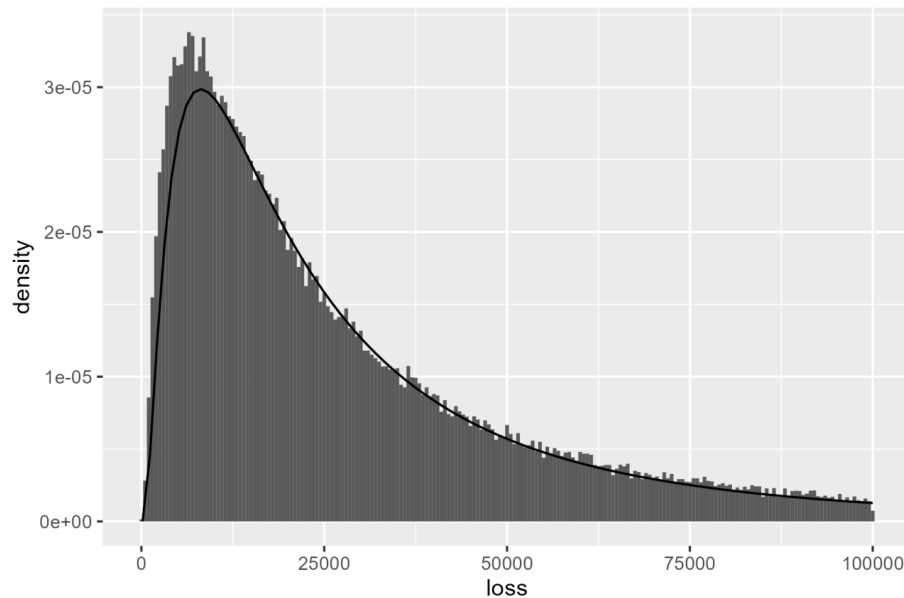
We can also  
update the  
color, if  
desired.



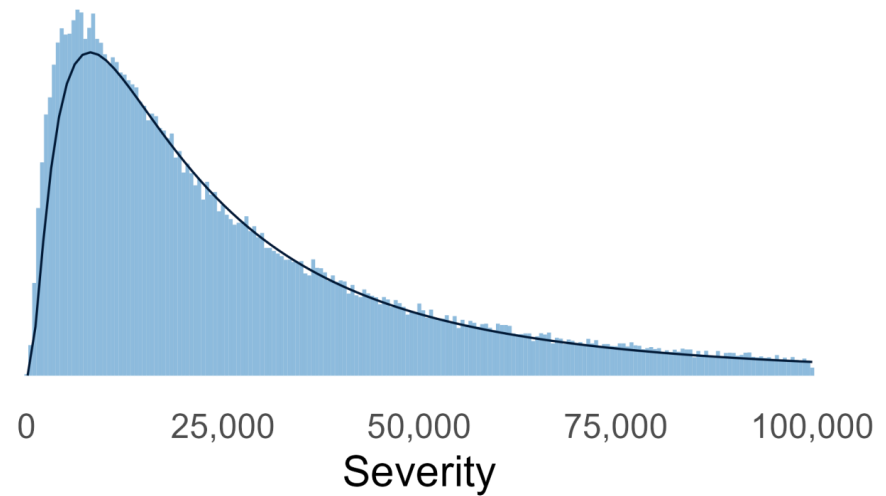
```
...  
geom_histogram(aes(x = loss, after_stat(density)),  
               fill = pal_CAS["light_blue"]) +  
stat_function(fun = dlnorm,  
             args = list(meanlog = 10, sdlog = 1.0),  
             color = pal_CAS["dark_blue"]) +  
...
```



# Comparison



Fitted Severity Curve

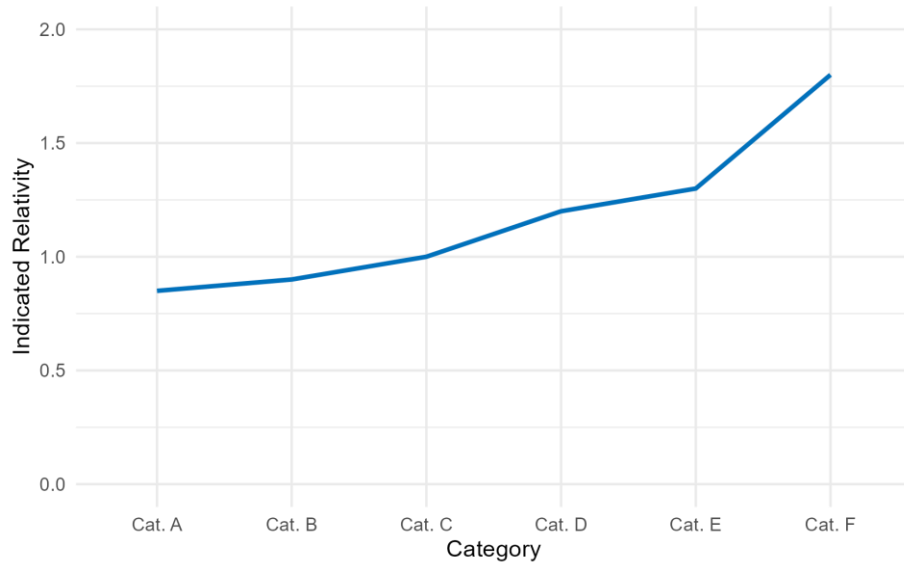


# APPENDIX

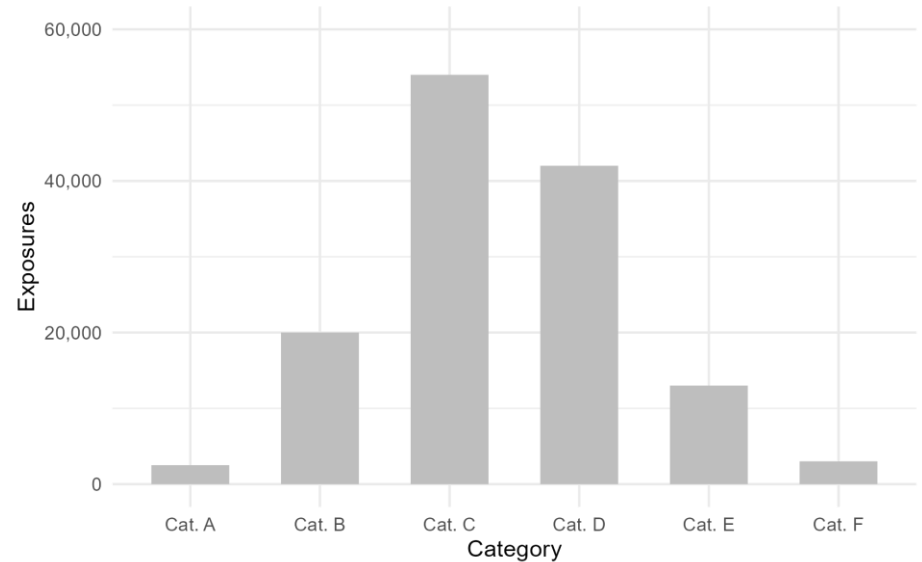


# Two Axes vs Two Charts

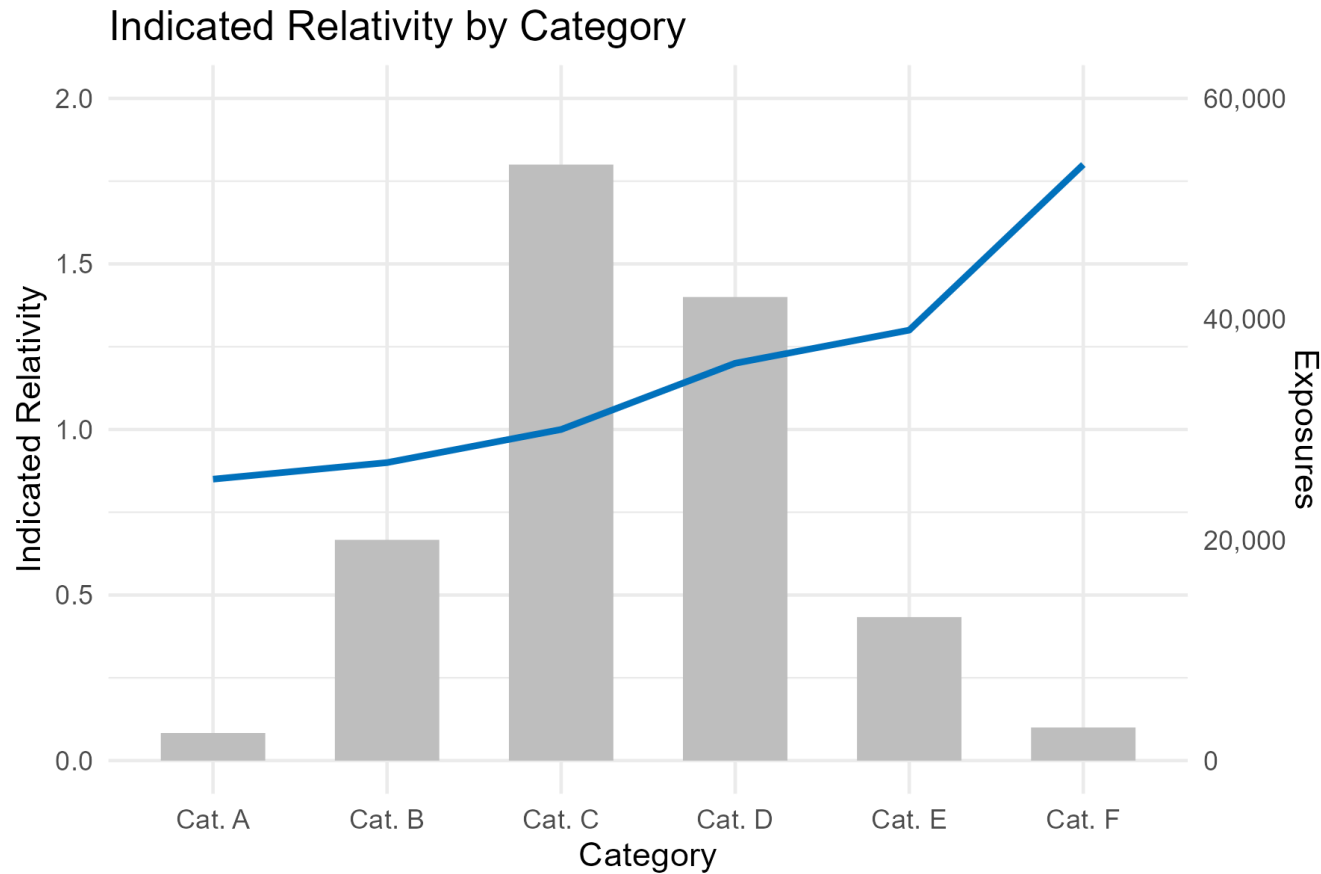
Indicated Relativity by Category



Indicated Relativity by Exposure Count & Category

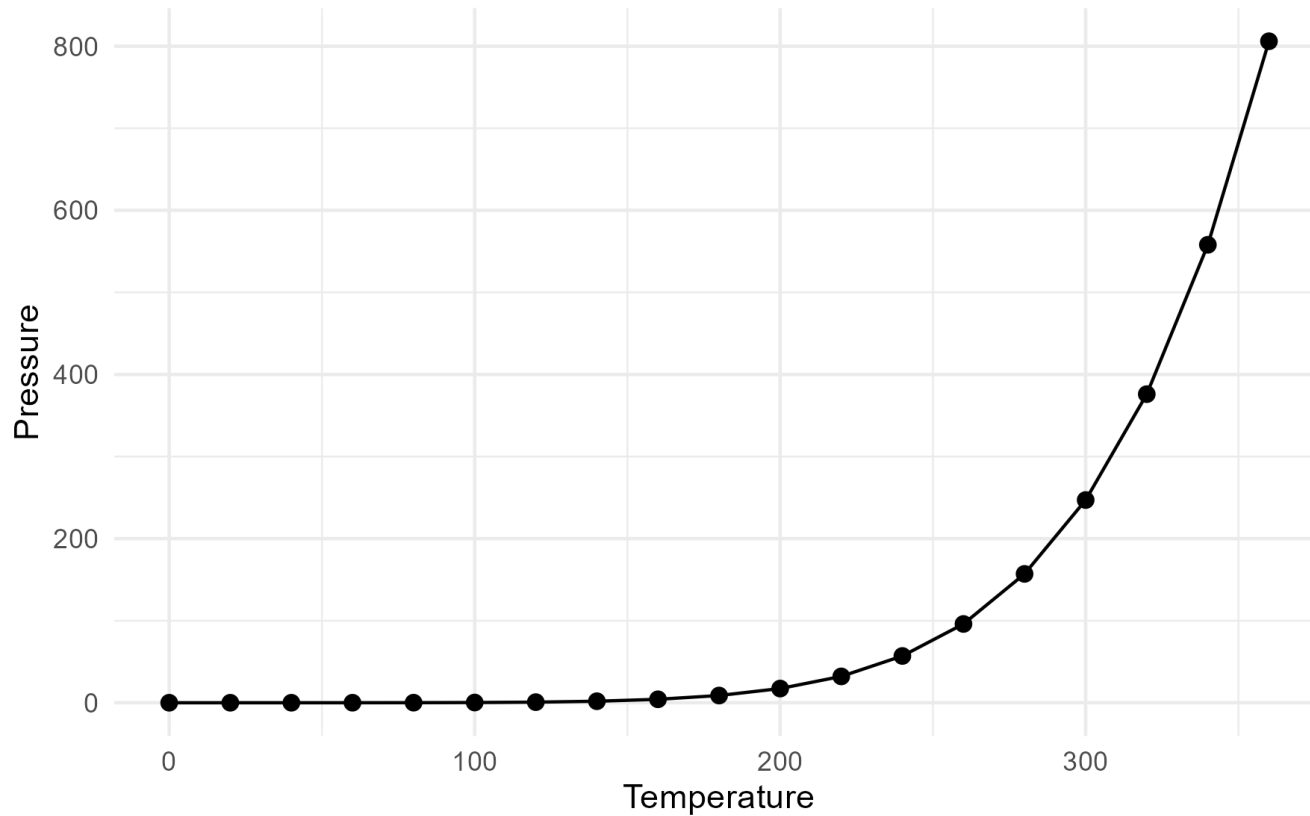


# Two Axes vs Two Charts



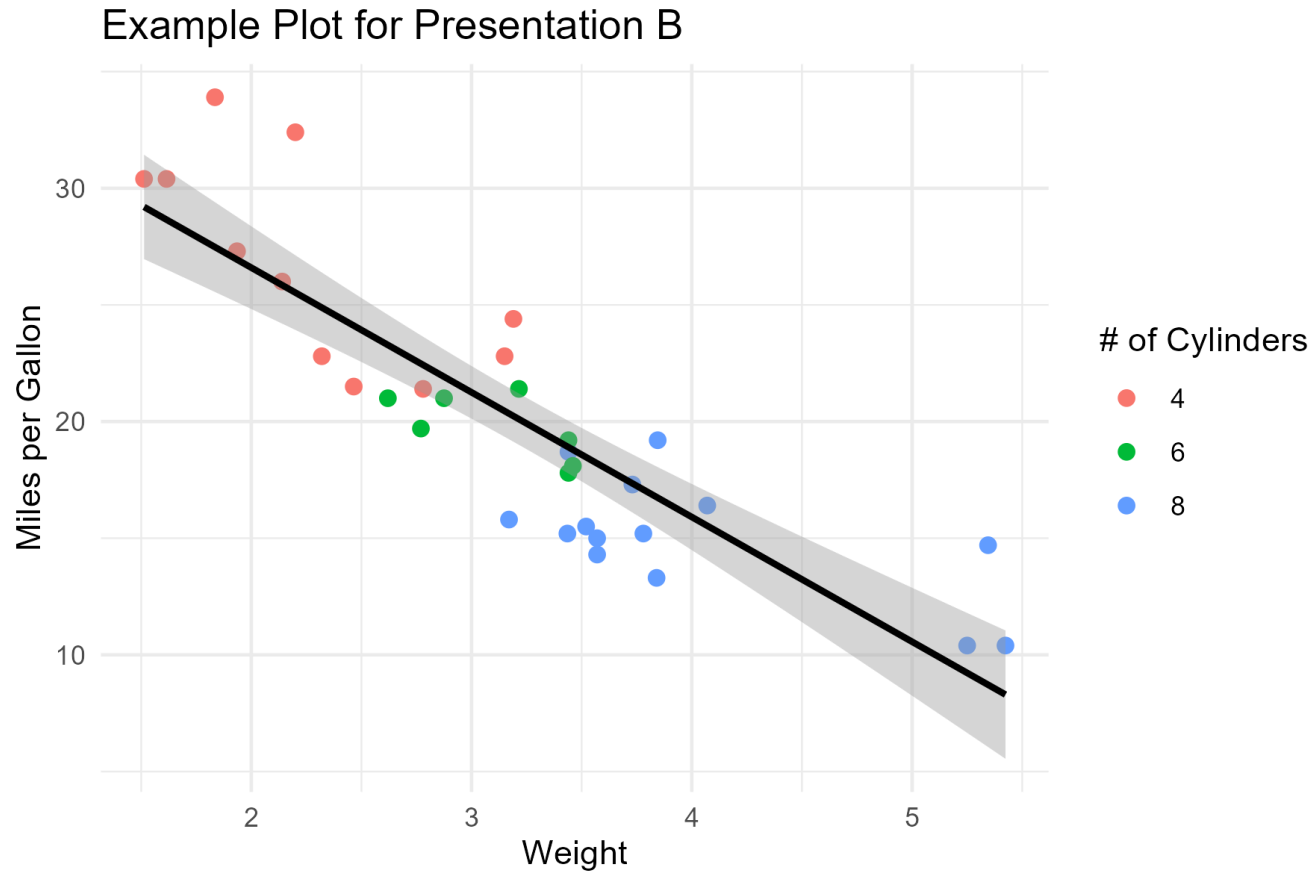
# Example Plot A

Example Plot for Presentation A





# Example Plot B



# Example Plot C

