

2018 Predictive Analytics Symposium

Session 15: ALL - Visualization: A Picture Speaks a Thousand Words

[SOA Antitrust Compliance Guidelines](#)
[SOA Presentation Disclaimer](#)

Every picture tells a story ... don't it?

Sam Castillo and Brian A. Fannin

September 19, 2019

Where to find this

This presentation may be found at:

https://pirategrunt.com/soa_symposium_2019/

Code to produce the examples and slides:

https://github.com/PirateGrunt/soa_symposium_2019

What we'll talk about

- Communication efficiency
- Every picture tells a story

Communication efficiency

Which of these two numbers is larger?

How about these two?

1011

1001

These?

IX

XI

And these?

9

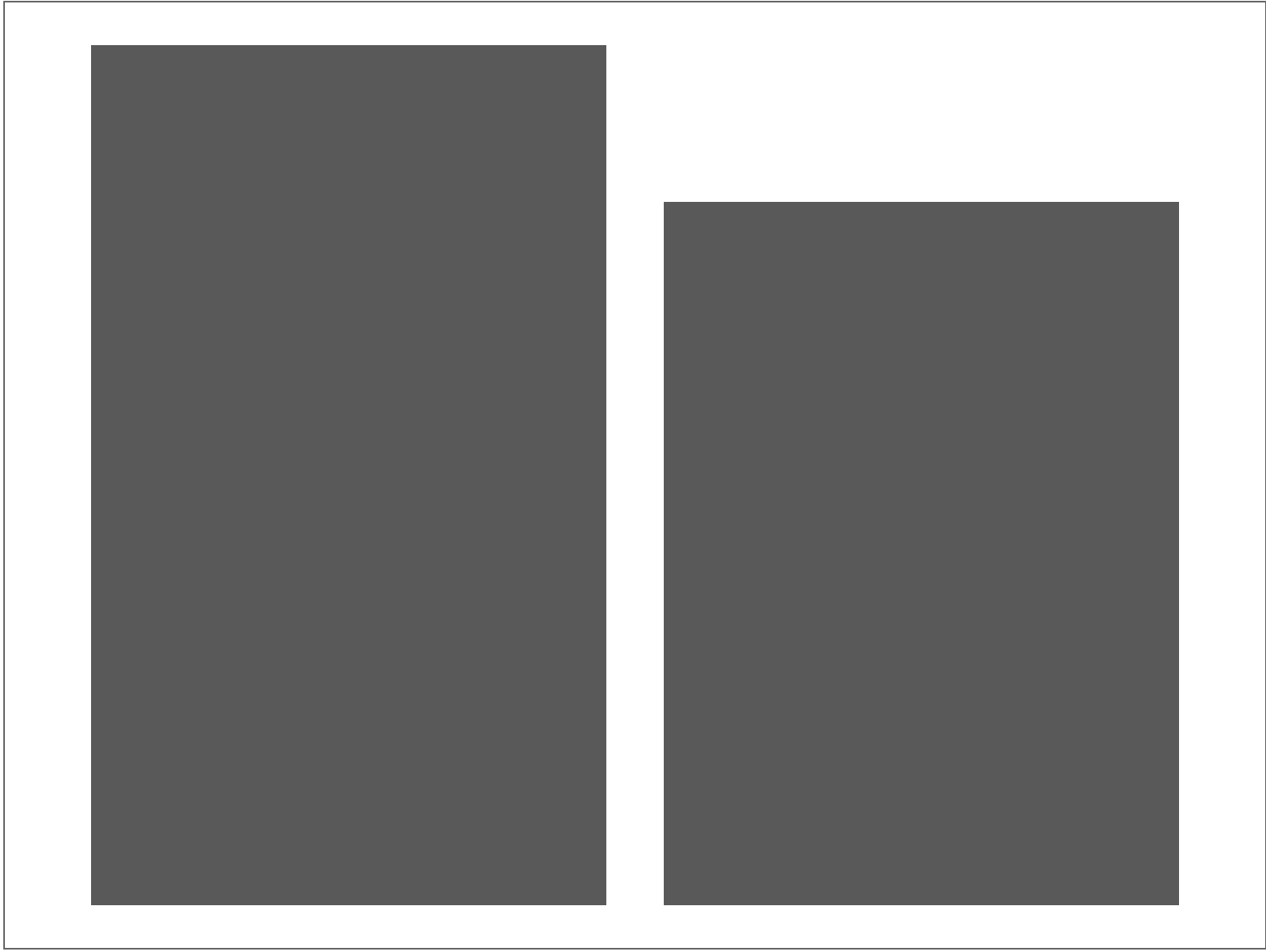
B

These?

十一

九

How about these?



nine

9

neun

1001

九

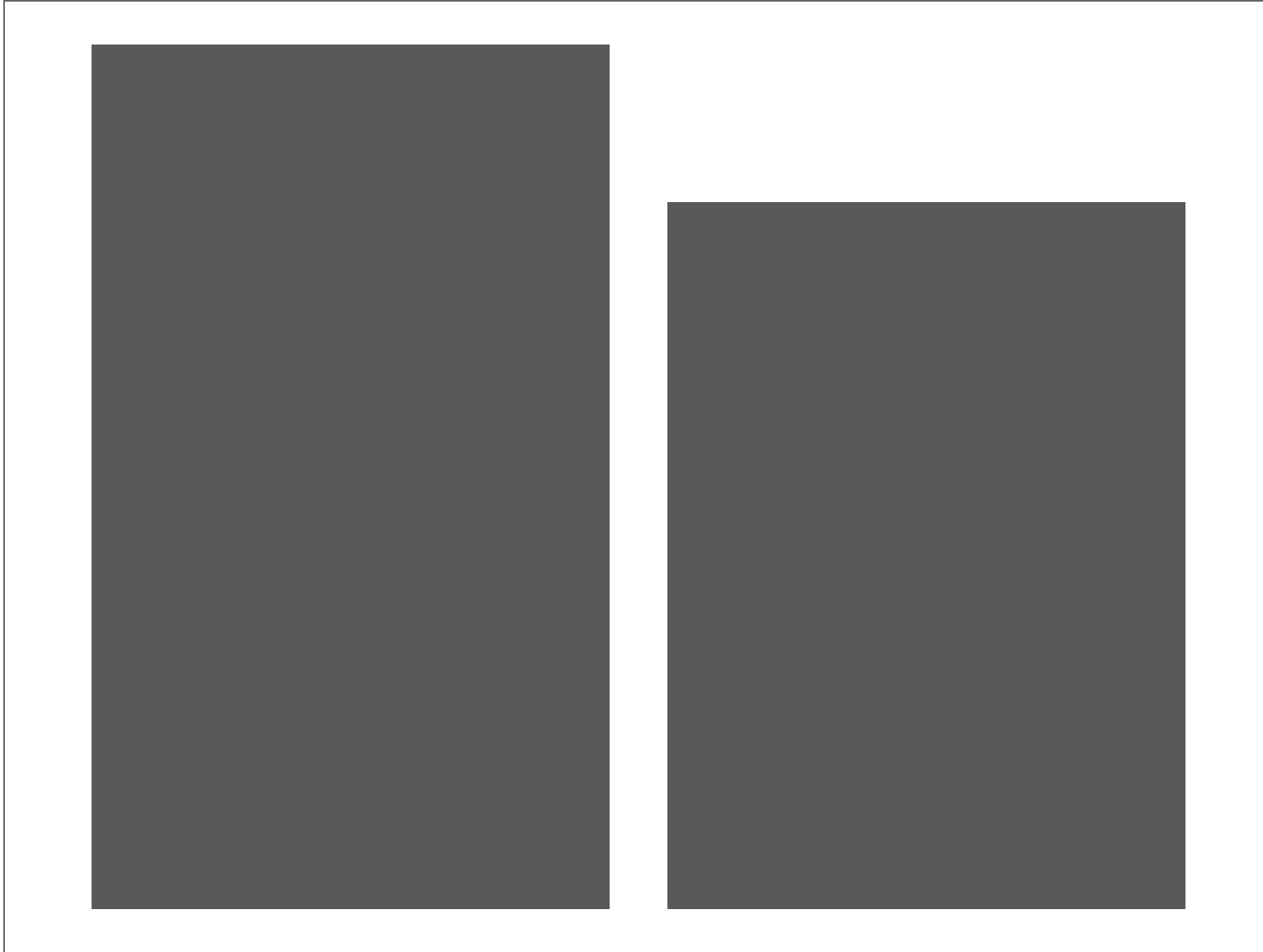
IX

nueve

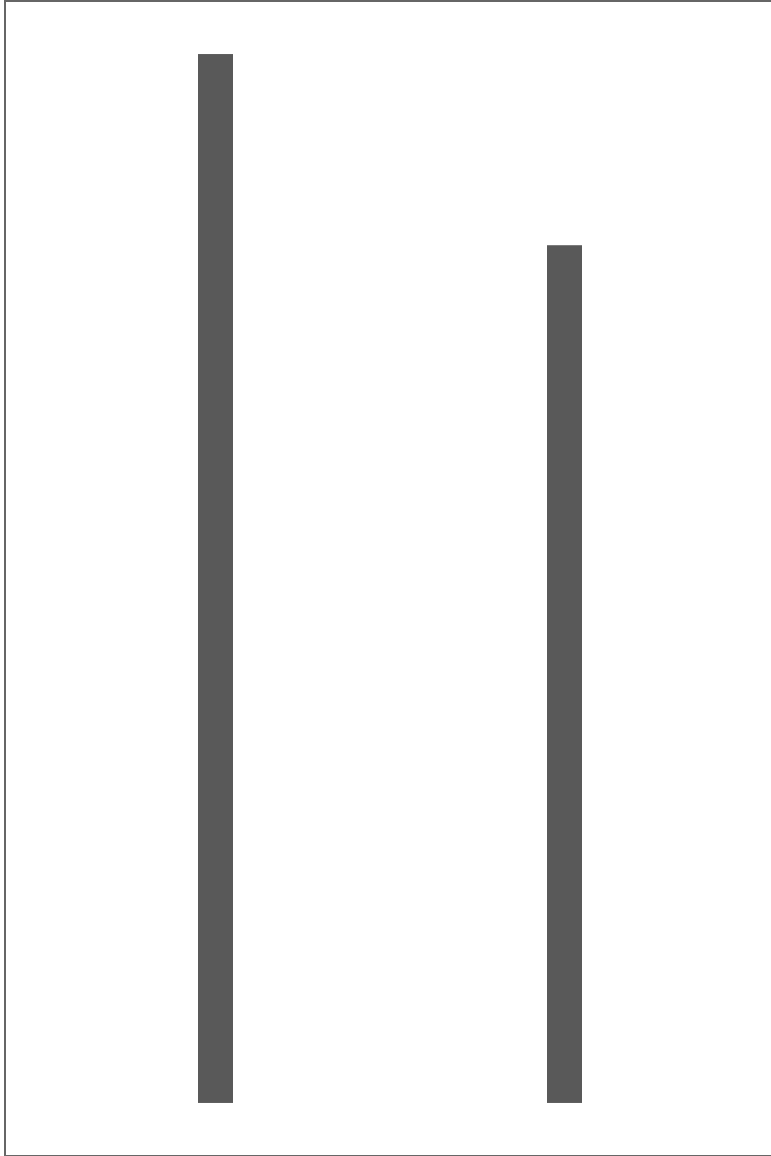
Arabic or sanskrit are no more legitimate than any other representation of numbers.

Be prepared to accept the idea that there are circumstances when geometric primitives may be understood *faster*.

This is actually *too much* information



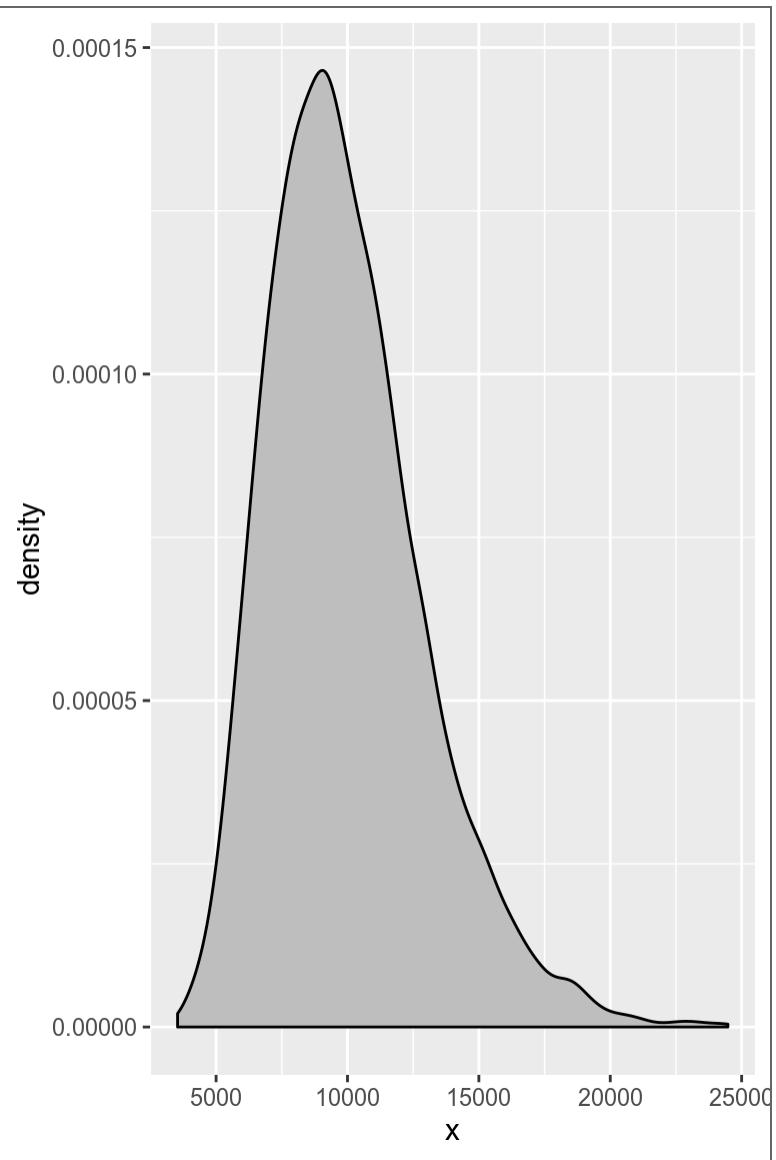
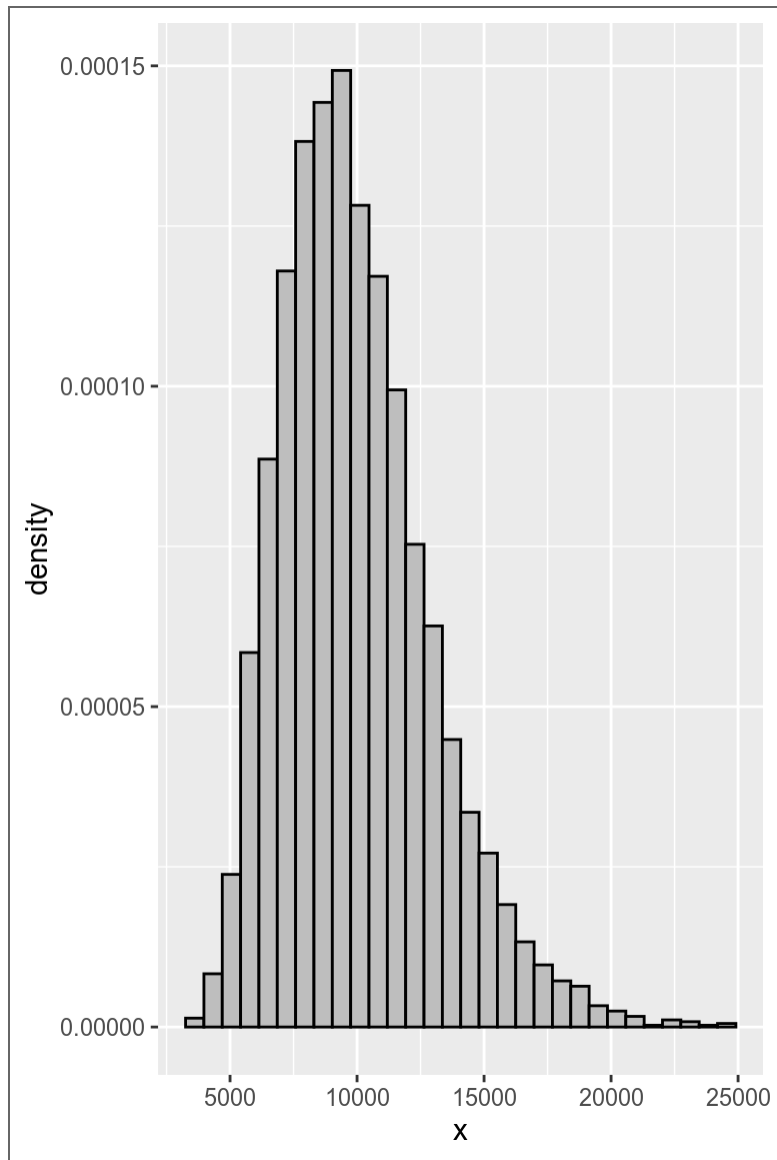
This is better

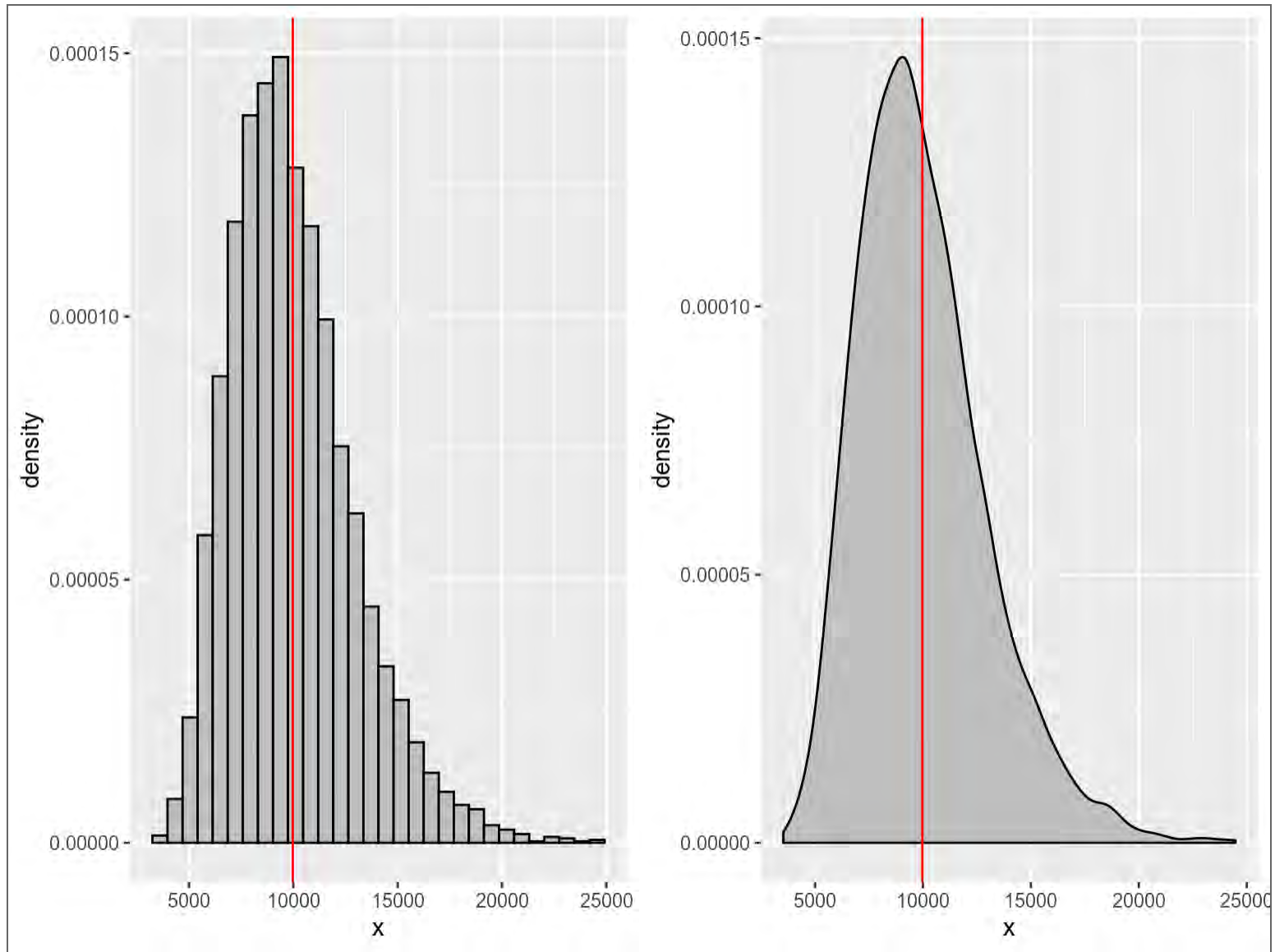


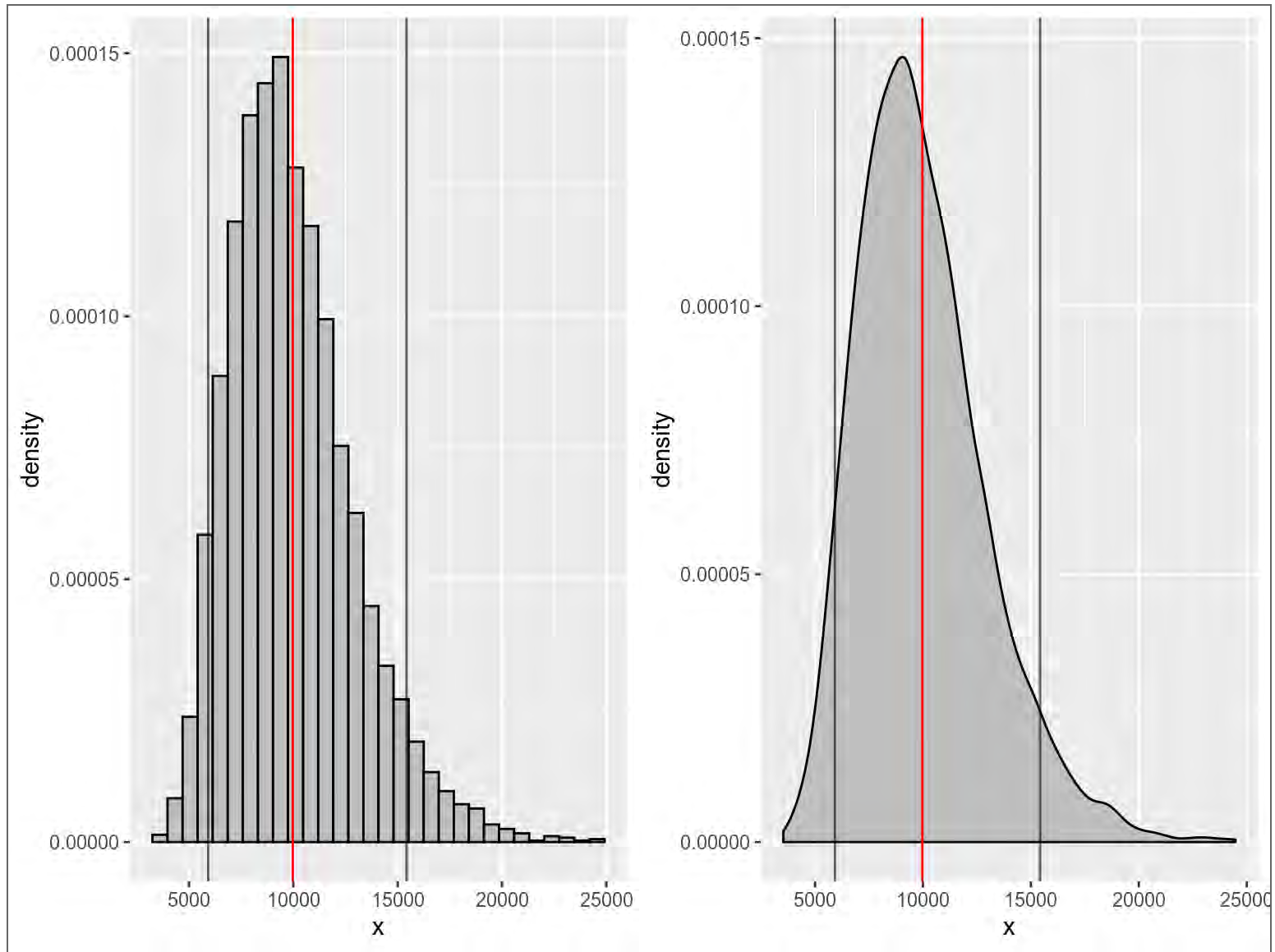
Statistics maps a set of many numbers into a set of fewer numbers.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3534	7854	9562	9973	11644	24476

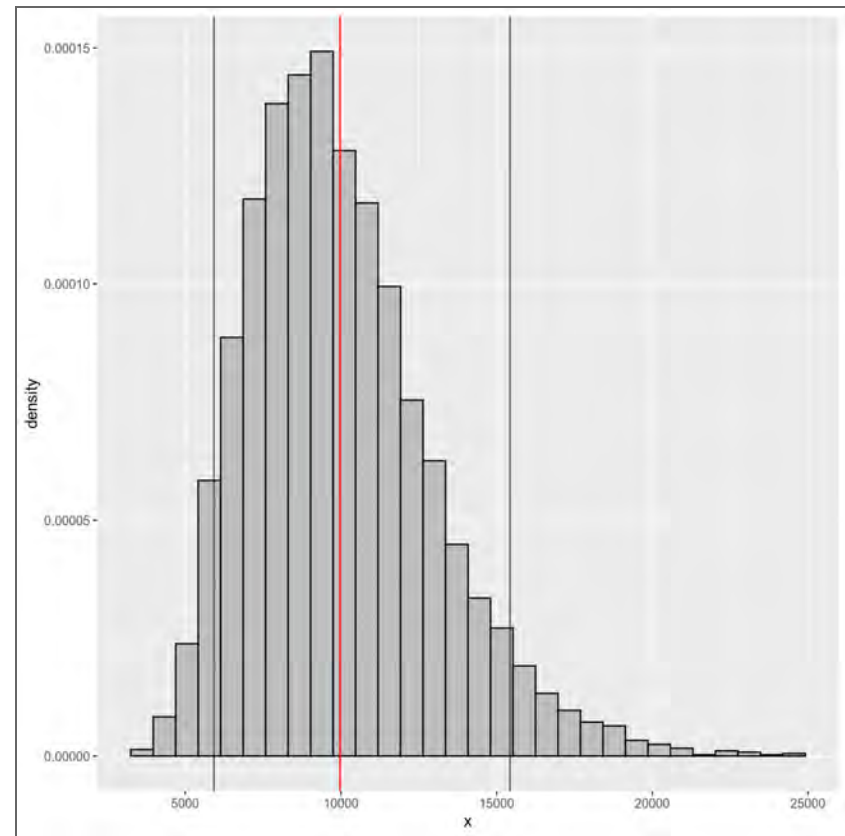
metric	x
Min.	3,534
1st Qu.	7,854
Median	9,562
Mean	9,973
3rd Qu.	11,644
Max.	24,476







metric	x
Min.	3,534
1st Qu.	7,854
Median	9,562
Mean	9,973
3rd Qu.	11,644
Max.	24,476



Summary statistics are *always* a reduction of information.

Visualization presents (almost) *all* of the data. The reductions are made with our eyes.

Potential complaint

Focus on visual design places undue emphasis on superfluous characteristics like color, font, etc.

What's the most important word in the text which follows?

The rate for territory X must be **increased** by 10.4%.

And this one?

The rate for **territory** X must be increased by 10.4%.

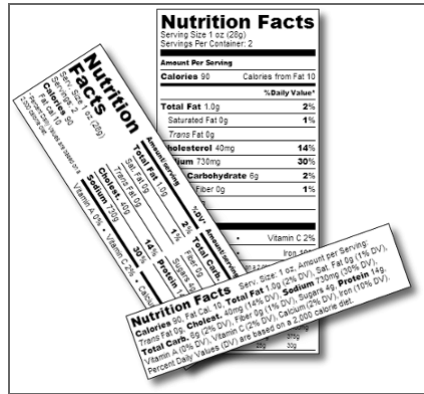
“Inessential” matters. Emphasis is an element of communication and therefore comprehension.

Are you ready to buy stock in this company?

This year we plan to build on last year's renewed profitability.

A picture can speak a thousand words

A real life example: improving my diet



Speed test: which is healthier, A or B?

Nutrition Facts	
Wheaties	
Serving Size: <input type="text" value="3"/> cup (108g)	
Amount Per Serving	
Calories 381	Calories from Fat 22.5
% Daily Value*	
Total Fat 2.5g	4%
Saturated Fat 0.5g	3%
Trans Fat 0g	
Polyunsaturated Fat 1.3g	
Monounsaturated Fat 0.4g	
Cholesterol 0mg	0%
Sodium 792mg	33%
Potassium 354mg	10%
Total Carbohydrates 90g	30%
Dietary Fiber 10.8g	43%
Sugars 16.5g	
Protein 9g	
Vitamin A	39%
Vitamin C	39%
Calcium	6%
Iron	201%

* Percent Daily Values are based on a 2000 calorie diet.

Nutrition Facts	
Frosted Flakes Cereal	
Serving Size: <input type="text" value="1"/> cup (42g)	
Amount Per Serving	
Calories 155	Calories from Fat 6.4
% Daily Value*	
Total Fat 0.7g	1%
Saturated Fat 0.2g	1%
Trans Fat 0g	
Polyunsaturated Fat 0.3g	
Monounsaturated Fat 0.1g	
Cholesterol 0mg	0%
Sodium 197mg	8%
Potassium 32mg	1%
Total Carbohydrates 37g	12%
Dietary Fiber 0.9g	4%
Sugars 15g	
Protein 1.7g	
Vitamin A	13%
Vitamin C	17%
Calcium	0.1%
Iron	62%

* Percent Daily Values are based on a 2000 calorie diet.

Weaknesses

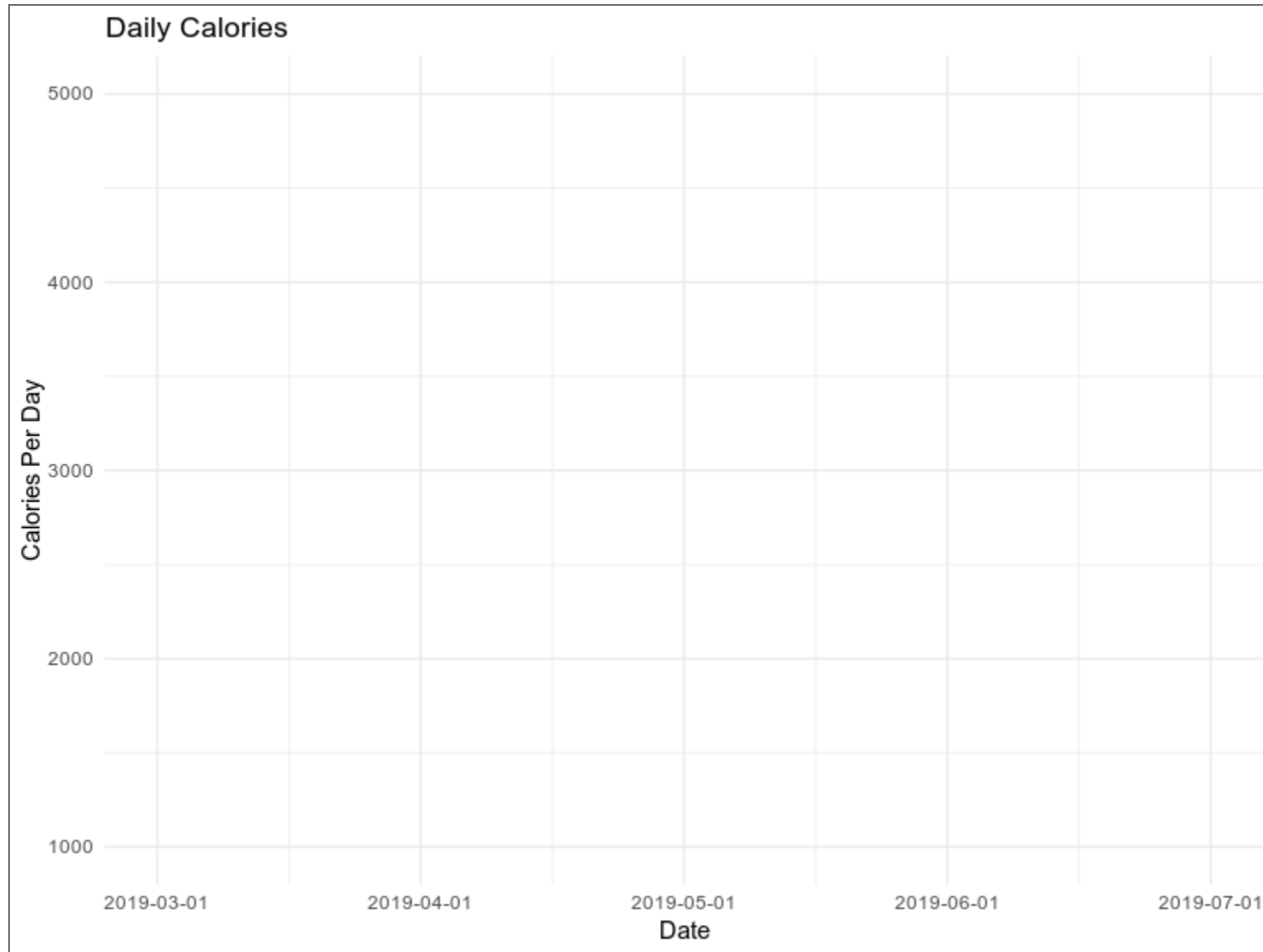
- Serving sizes are inconsistent
- Converting units requires a calculator
- Each food has it's own label
- Data collection is slow

A better data source

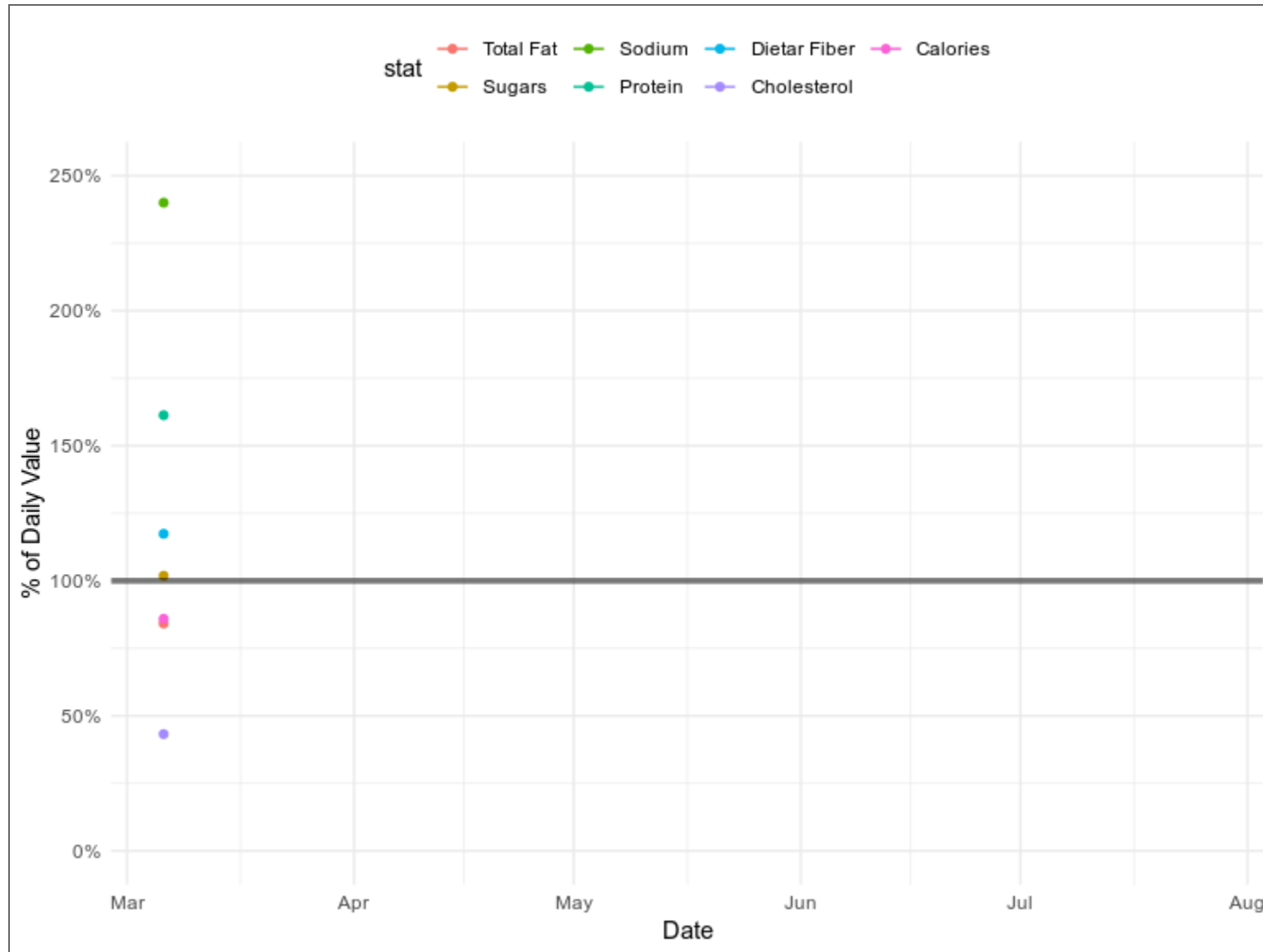
Your Orders

Order placed	Items	Total	Rating	
Aug 10, 2019, 8:59 AM	16	\$84.12	Not rated	View order detail
ALDI				
				Add all items to cart
				Rate & Tip
Order placed	Items	Total	Rating	
Jul 27, 2019, 3:02 PM	21	\$97.86	Not rated	View order detail
ALDI				
				Add all items to cart
				Rate & Tip
Order placed	Items	Total	Rating	
Jul 20, 2019, 11:17 AM	15	\$64.64	Not rated	View order detail
ALDI				
				Add all items to cart
				Rate & Tip

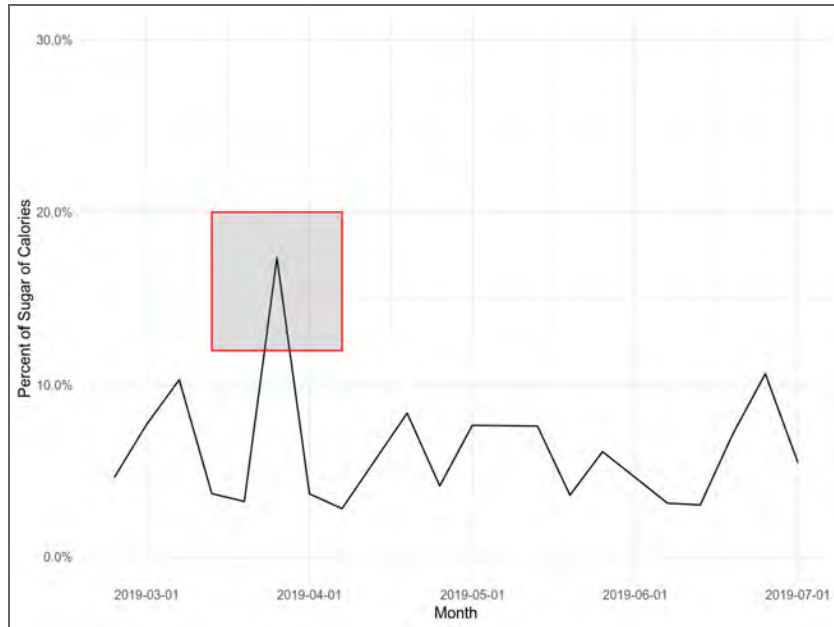
My daily calories for the last four months



The picture that speaks a thousand words



The price of triple chocolate cake



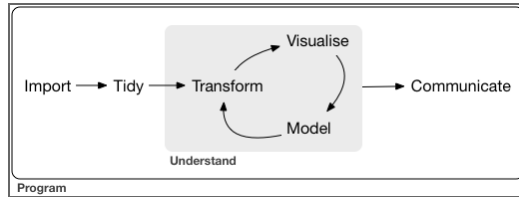
Triple Chocolate Cake

Nutrition Facts	value
Calories	22,000
Total Fat (g)	1,100
Saturated Fat (g)	200
Trans Fat (g)	0
Cholesterol (mg)	4,000
Sodium (mg)	20,000
Total Carbs (g)	2,900
Dietar Fiber (g)	50
Sugars (g)	2,000

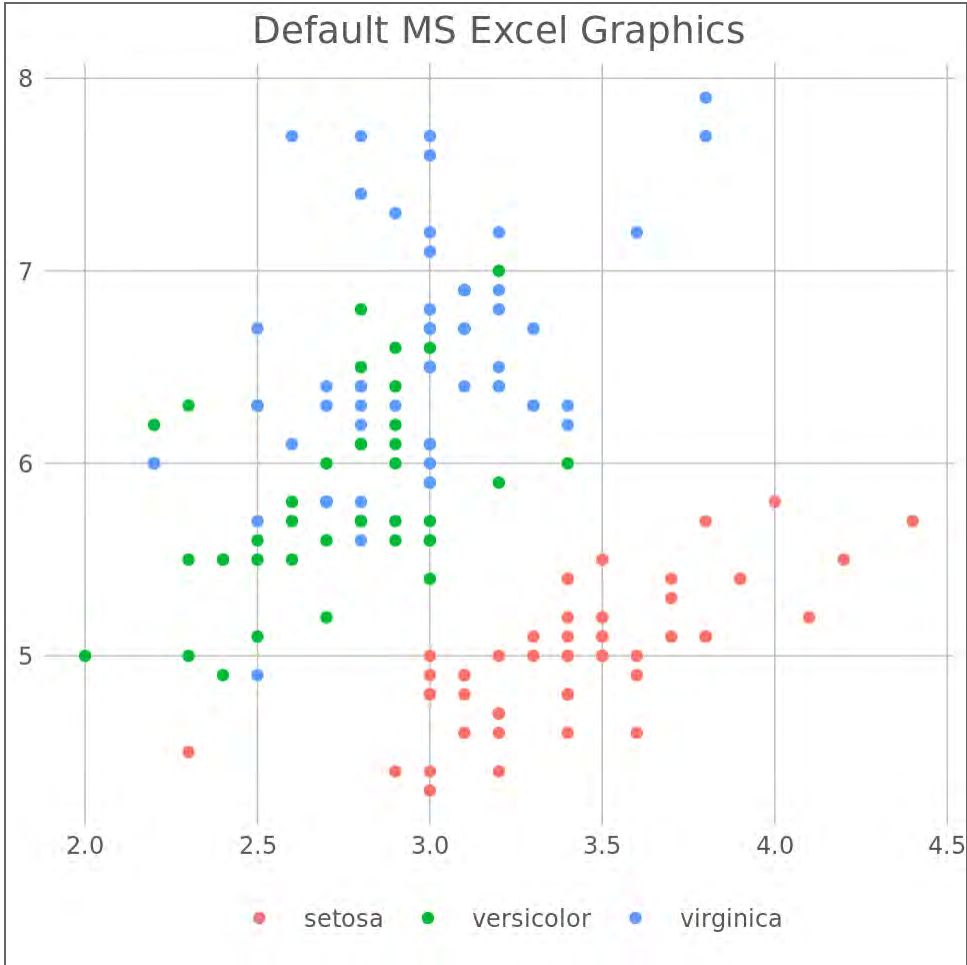
The two use cases for data visualization

- Exploratory analysis
- Persuading an audience

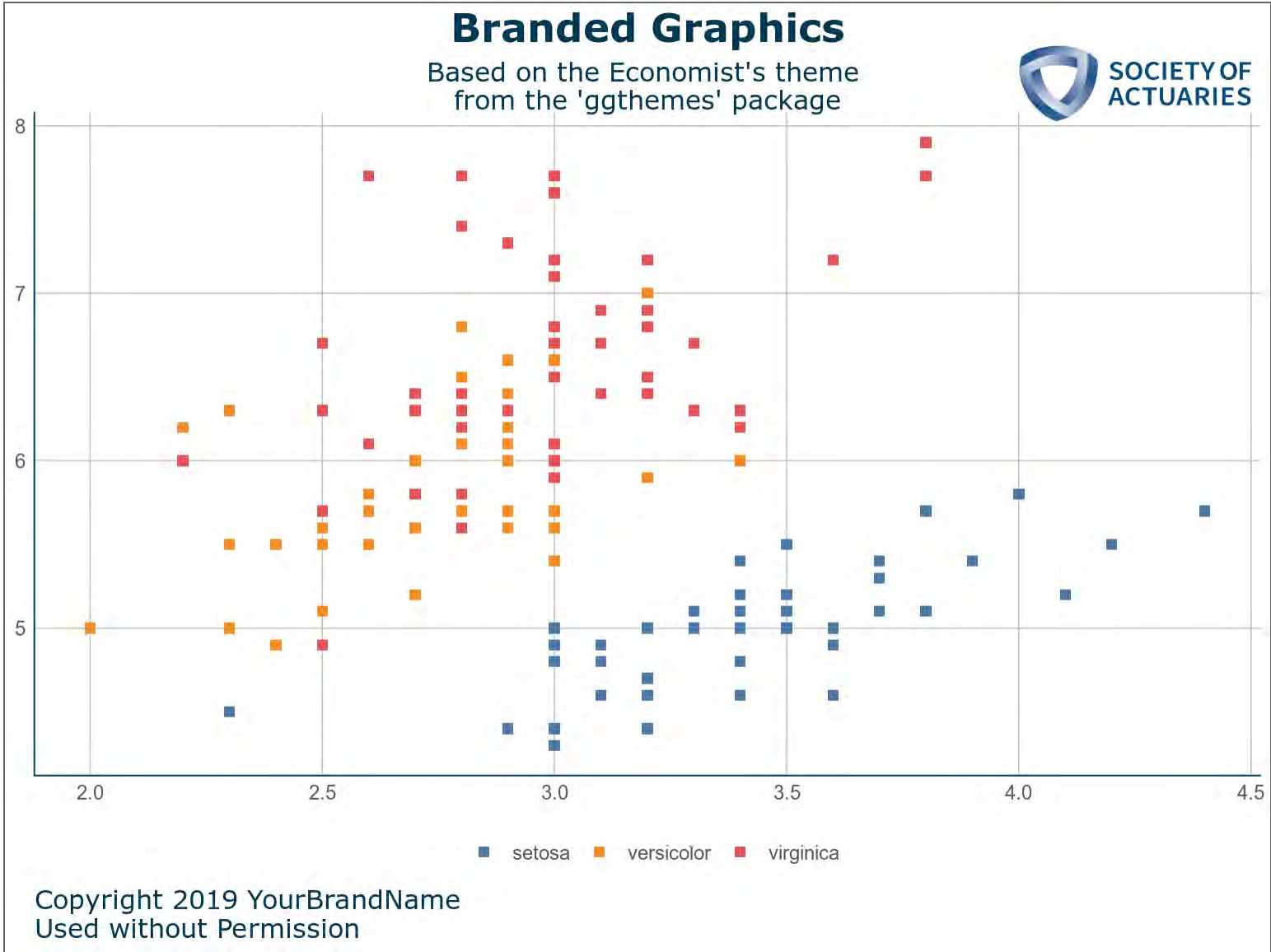
Exploratory analysis requires speed and iteration



Data visualization communicates your brand



Data visualization communicates your brand

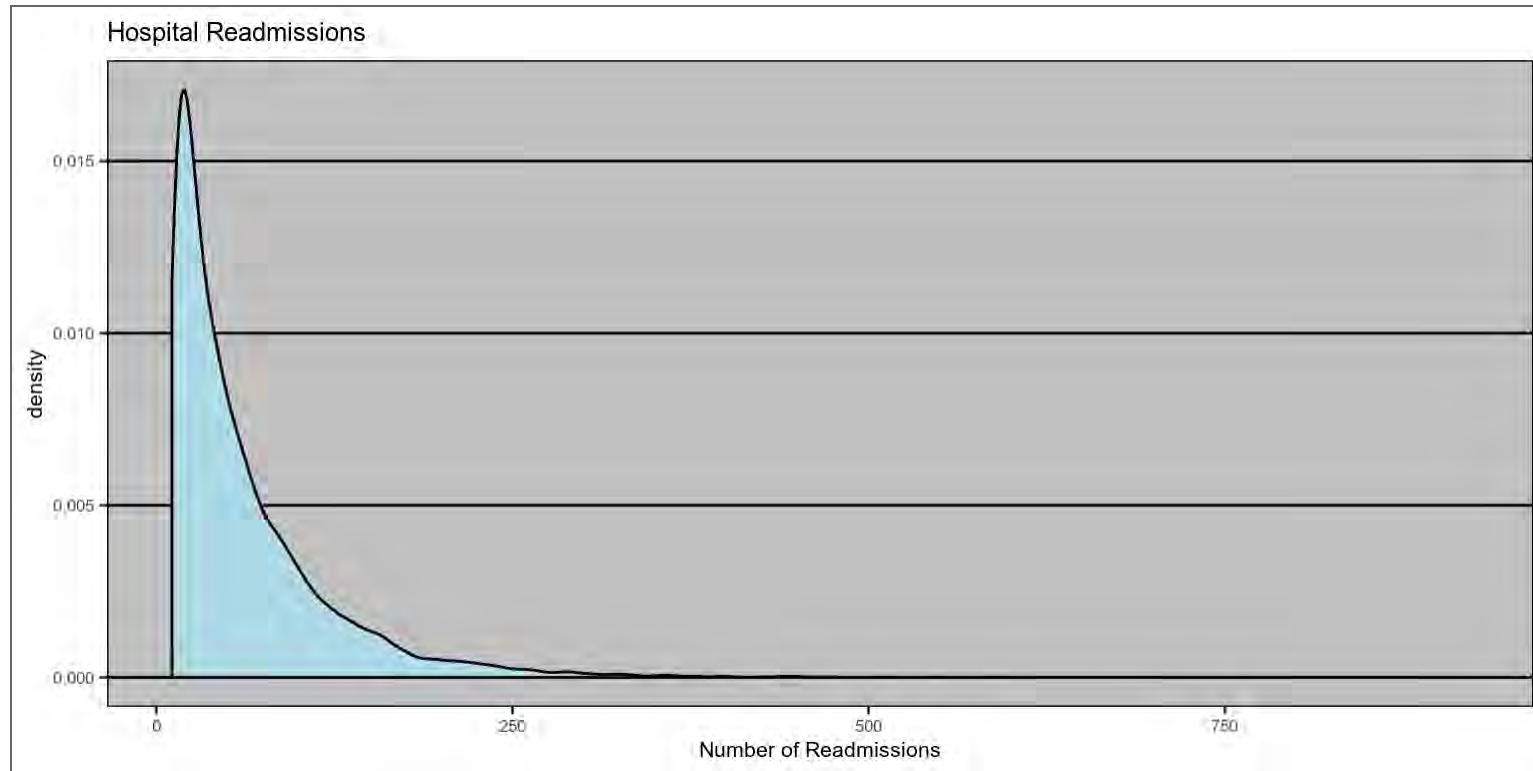


Automation is the key to speed

```
library(custom_library)
```

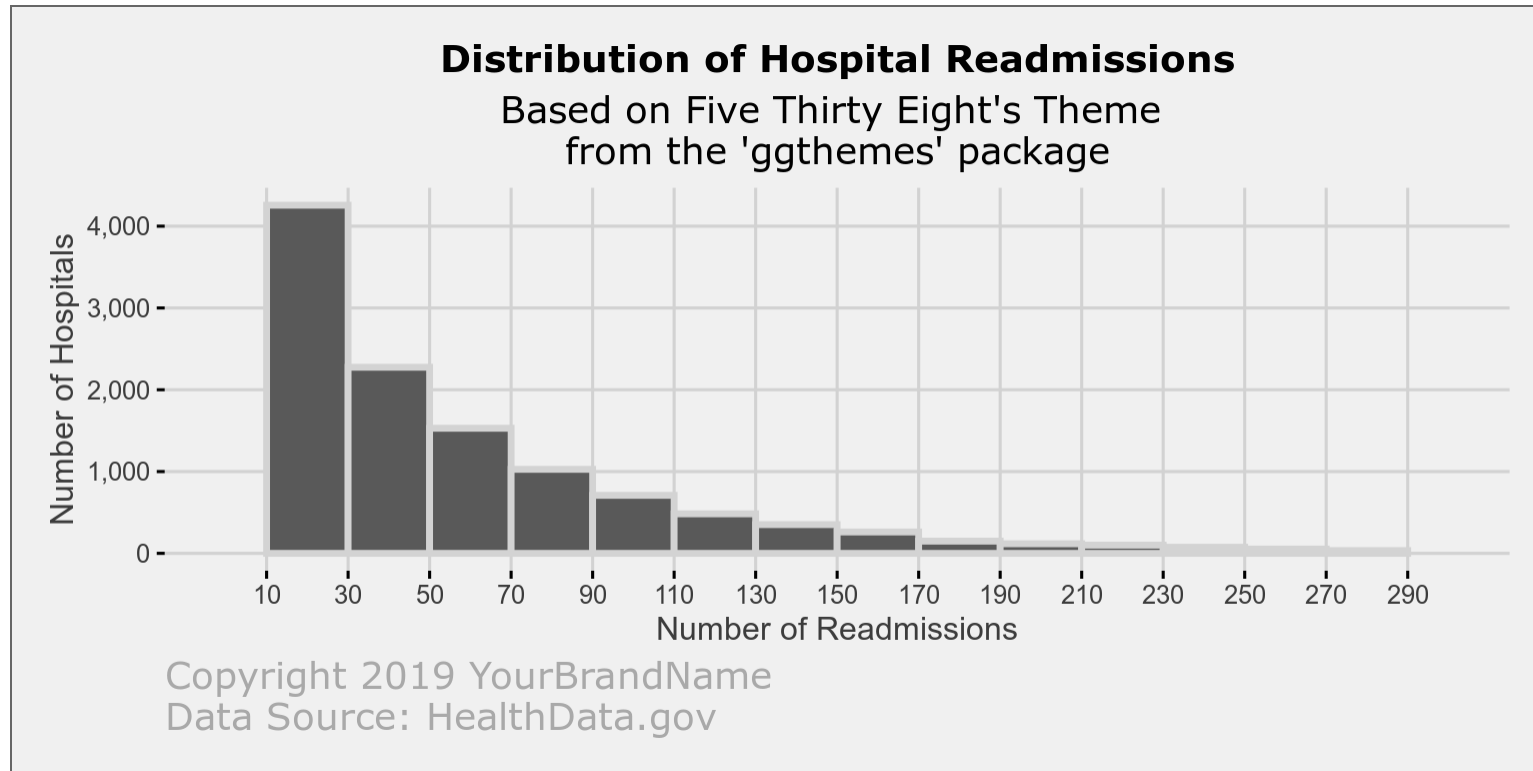
```
data %>%  
  ggplot(x, y) +  
  geom_point() +  
  my_custom_theme()
```

The best graphs are easy to read



- What are the y-axis “density” units?
- How can this be translated into english?

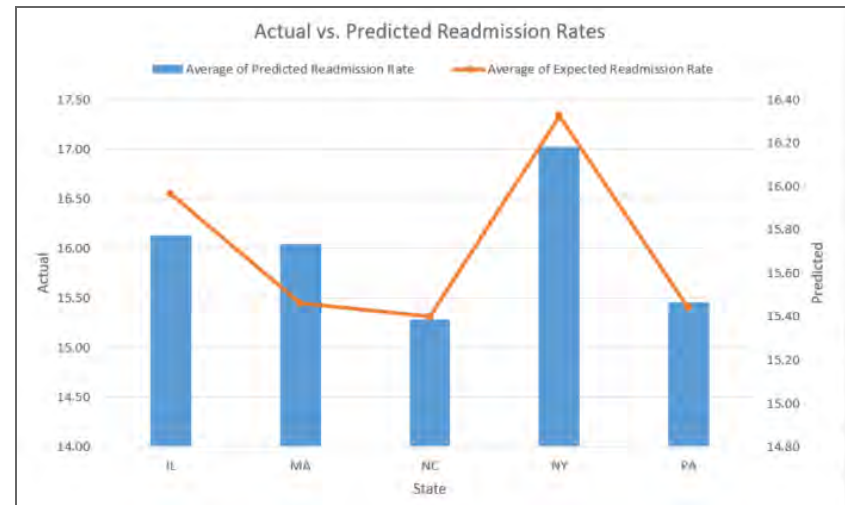
The best graphs are easy to read



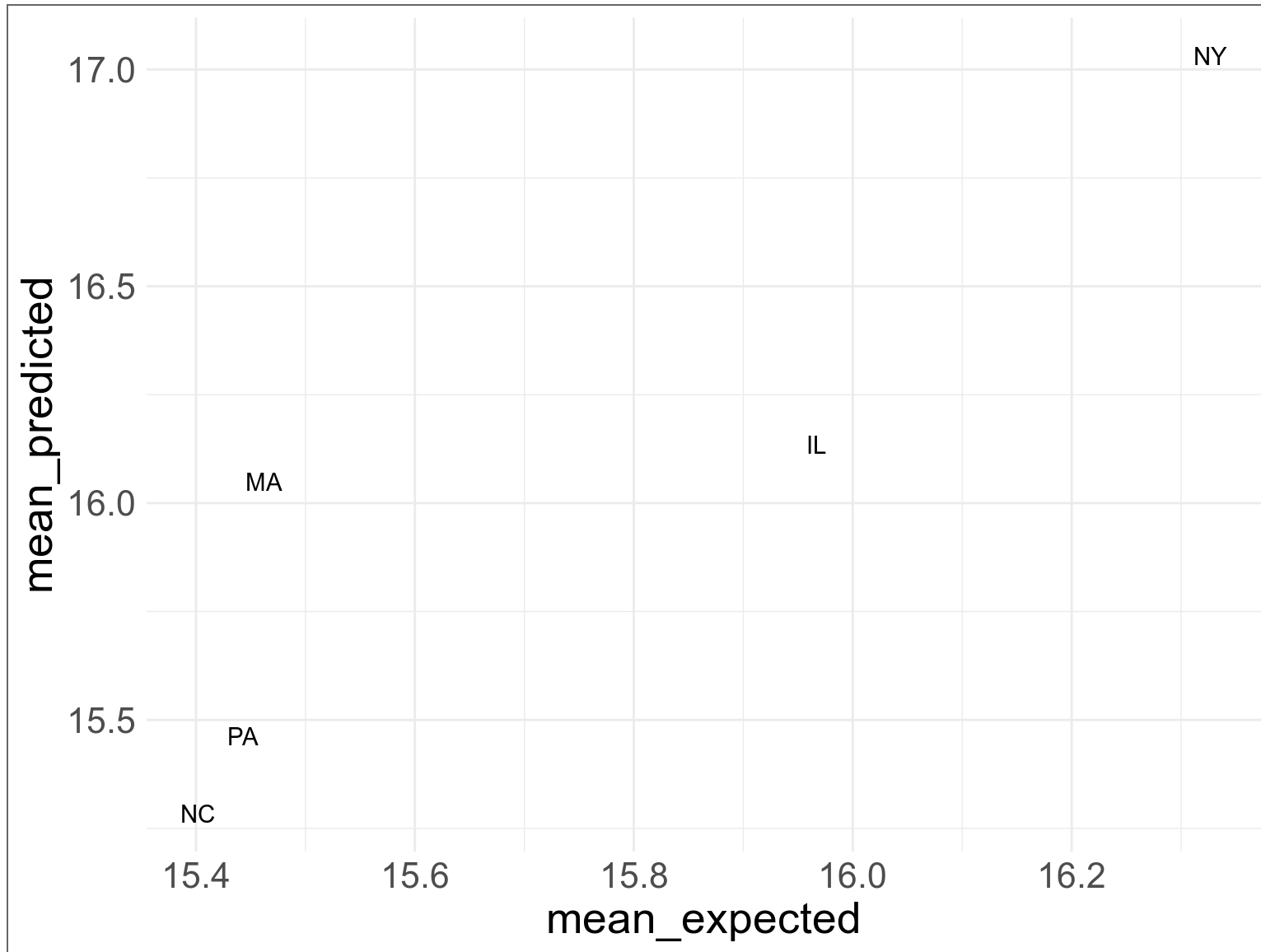
- In english: “There are just over 2,000 hospitals with between 30 and 50 readmissions”

Graphs should be unambiguous

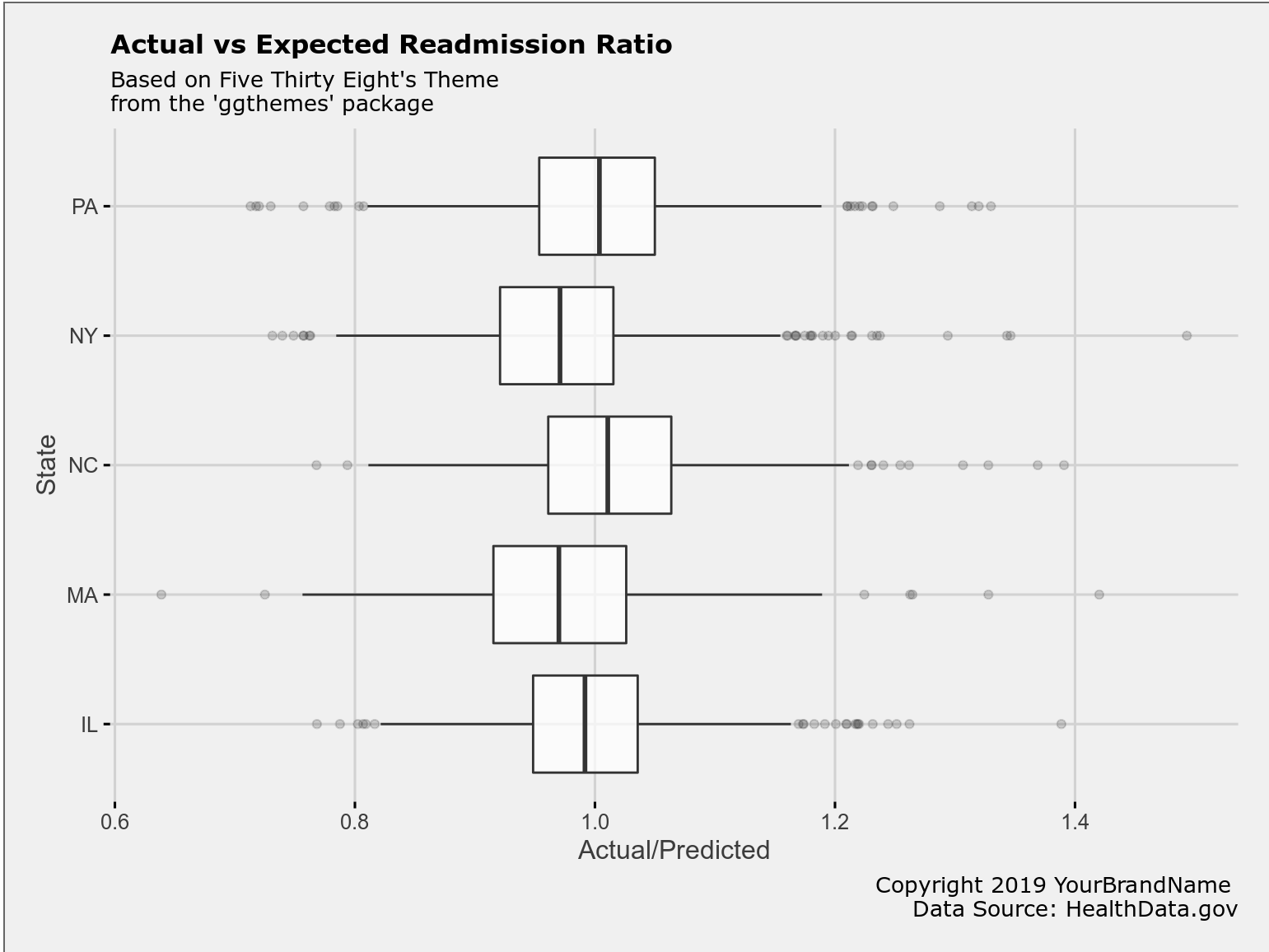
- Which axis is the Actual and which is the Predicted?
- The blue bars take up more visual space than the lines, but both units are equally important
- X-axis ontinuous scale does not make sense



A step in the right direction



Uncertainty should be shown



Persuading an audience is a *very* different challenge

How to Win Friends and Influence People - Dale Carnegie

1. You can't win an argument
2. Avoid making enemies - Never tell the the audience that they are wrong
3. Put yourself in the other person's shoes
4. Admit when you are wrong
5. Be friendly
6. Get the other person saying "yes, yes" immediately
7. Let the other person see the pattern and come to the conclusion themselves

Summary

For exploratory analysis

1. Understand the story behind the data
2. Focus on speed and flexibility

For persuading an audience, publishing, or selling your ideas

1. Build templates
2. Strive for simplicity and professionalism

Thank you!

Where to find this

This presentation may be found at:

https://pirategrunt.com/soa_symposium_2019/

Code to produce the examples and slides:

https://github.com/PirateGrunt/soa_symposium_2019