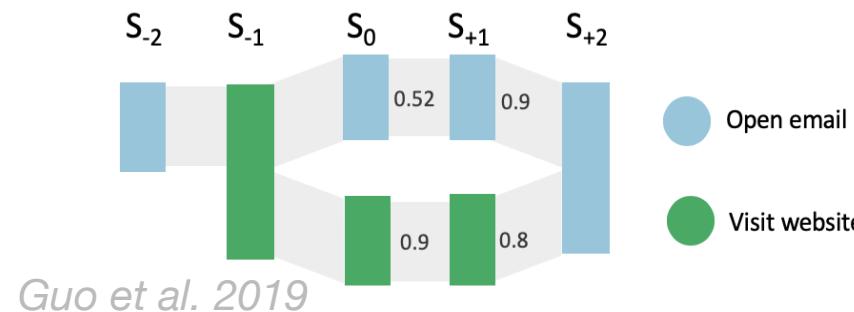


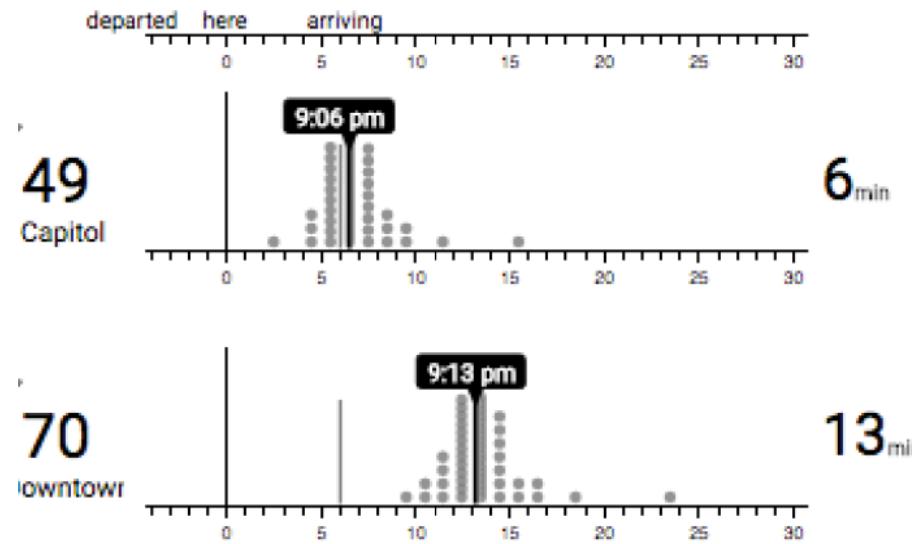
# A Probabilistic Grammar of Graphics

Xiaoying Pu  
Prelim presentation

## Machine learning

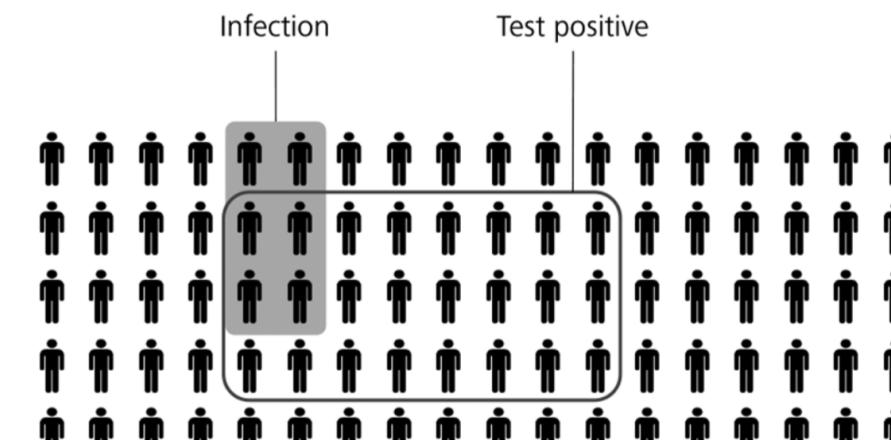


## Bus arrival time



(Fernandes et al. 2018)

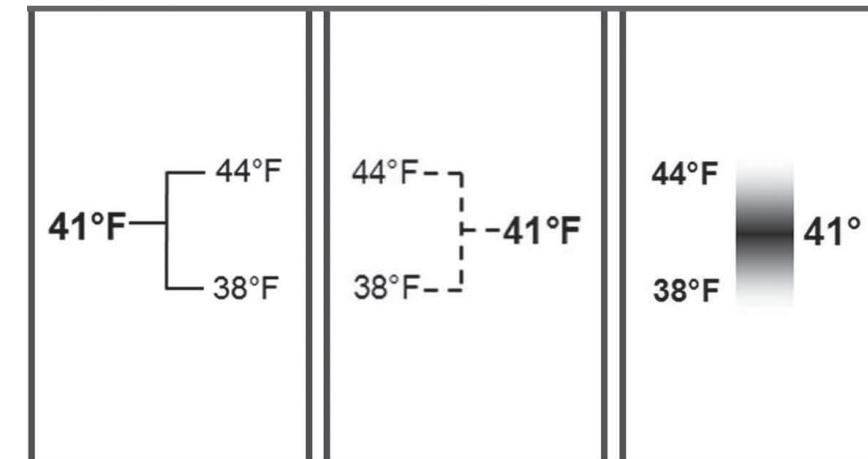
## Medical risk communication



(Binder, Krauss, and Bruckmaier 2015)

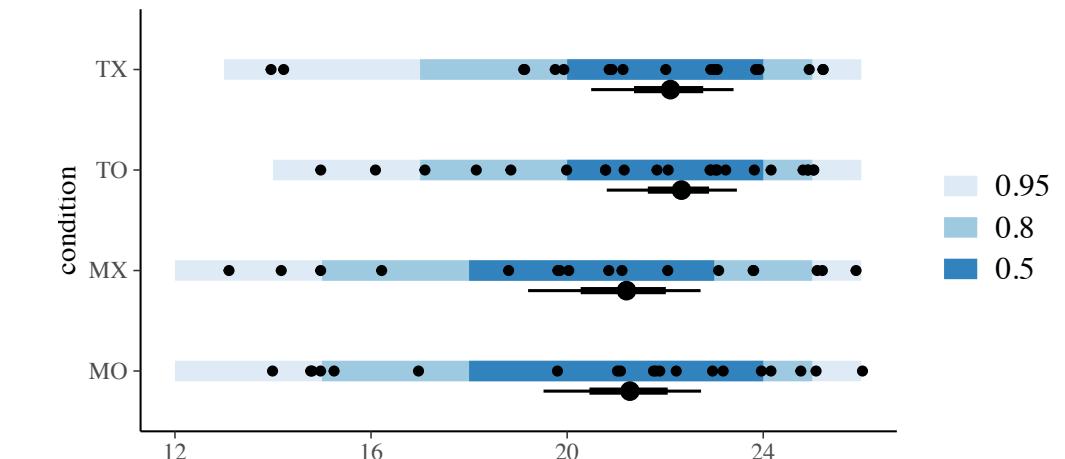
# Probabilistic visualizations: same substrate, many domains

## Weather forecast



(Joslyn and LeClerc 2013)

## Statistical modeling



# What could possibly go wrong?

|                   | mpg  | cyl | am |
|-------------------|------|-----|----|
| Mazda RX4         | 21.0 | 6   | 1  |
| Mazda RX4 Wag     | 21.0 | 6   | 1  |
| Datsun 710        | 22.8 | 4   | 1  |
| Hornet 4 Drive    | 21.4 | 6   | 0  |
| Hornet Sportabout | 18.7 | 8   | 0  |
| Valiant           | 18.1 | 6   | 0  |

*What's the distribution of mileage  
and cylinders?*

--- a user's mental process

# What could possibly go wrong?

|                   | mpg  | cyl | am |
|-------------------|------|-----|----|
| Mazda RX4         | 21.0 | 6   | 1  |
| Mazda RX4 Wag     | 21.0 | 6   | 1  |
| Datsun 710        | 22.8 | 4   | 1  |
| Hornet 4 Drive    | 21.4 | 6   | 0  |
| Hornet Sportabout | 18.7 | 8   | 0  |
| Valiant           | 18.1 | 6   | 0  |

```
ggplot(mtcars) +  
  geom_density(aes(  
    x = mpg,
```

# What could possibly go wrong?

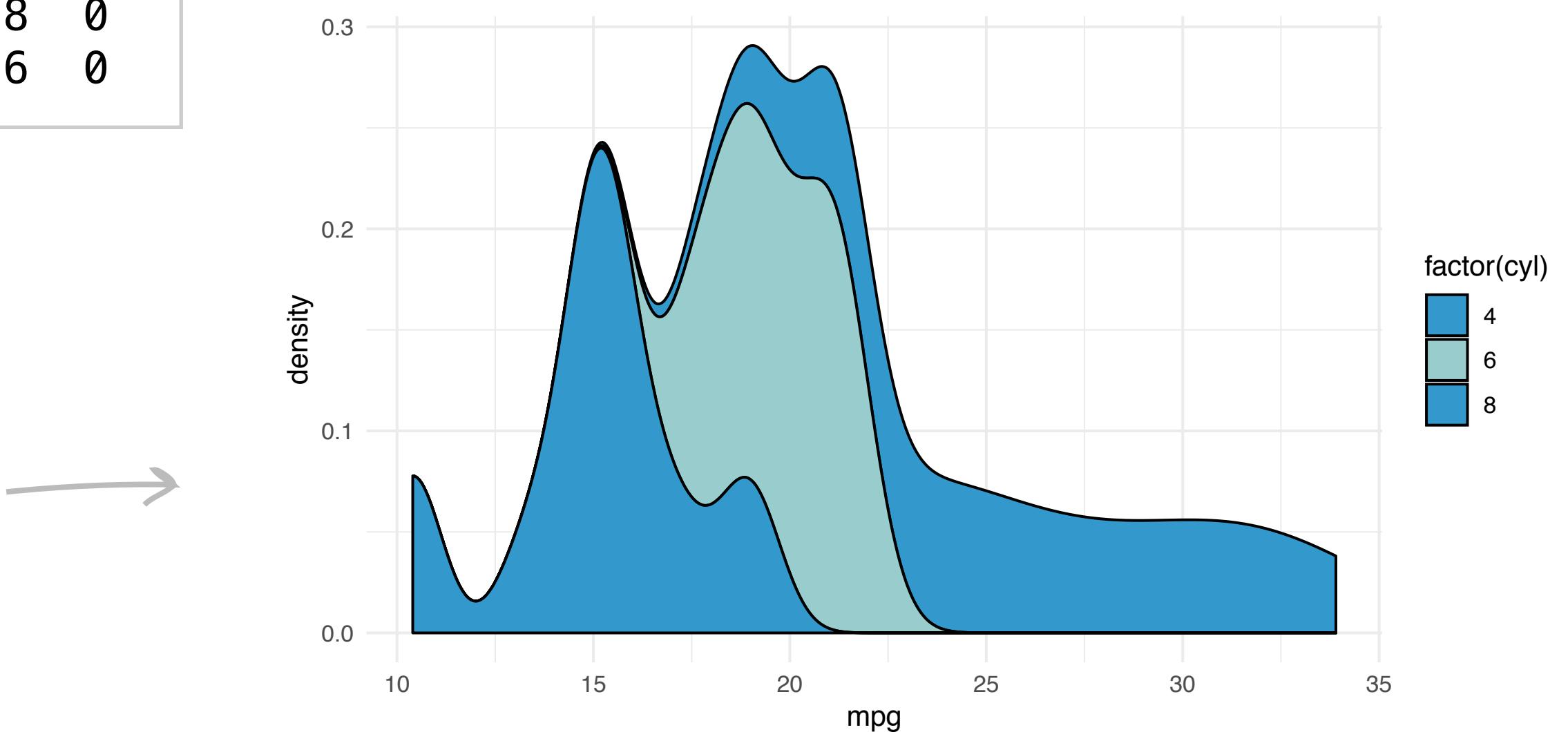
|                   | mpg  | cyl | am |
|-------------------|------|-----|----|
| Mazda RX4         | 21.0 | 6   | 1  |
| Mazda RX4 Wag     | 21.0 | 6   | 1  |
| Datsun 710        | 22.8 | 4   | 1  |
| Hornet 4 Drive    | 21.4 | 6   | 0  |
| Hornet Sportabout | 18.7 | 8   | 0  |
| Valiant           | 18.1 | 6   | 0  |

```
ggplot(mtcars) +  
  geom_density(aes(  
    x = mpg,  
    fill = cyl),  
  position = "stack")
```

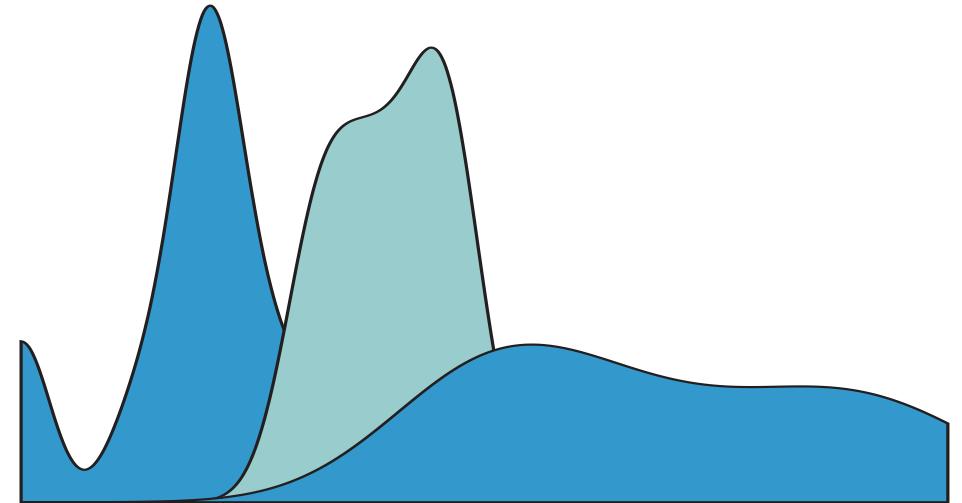
# What could possibly go wrong?

|                   | mpg  | cyl | am |
|-------------------|------|-----|----|
| Mazda RX4         | 21.0 | 6   | 1  |
| Mazda RX4 Wag     | 21.0 | 6   | 1  |
| Datsun 710        | 22.8 | 4   | 1  |
| Hornet 4 Drive    | 21.4 | 6   | 0  |
| Hornet Sportabout | 18.7 | 8   | 0  |
| Valiant           | 18.1 | 6   | 0  |

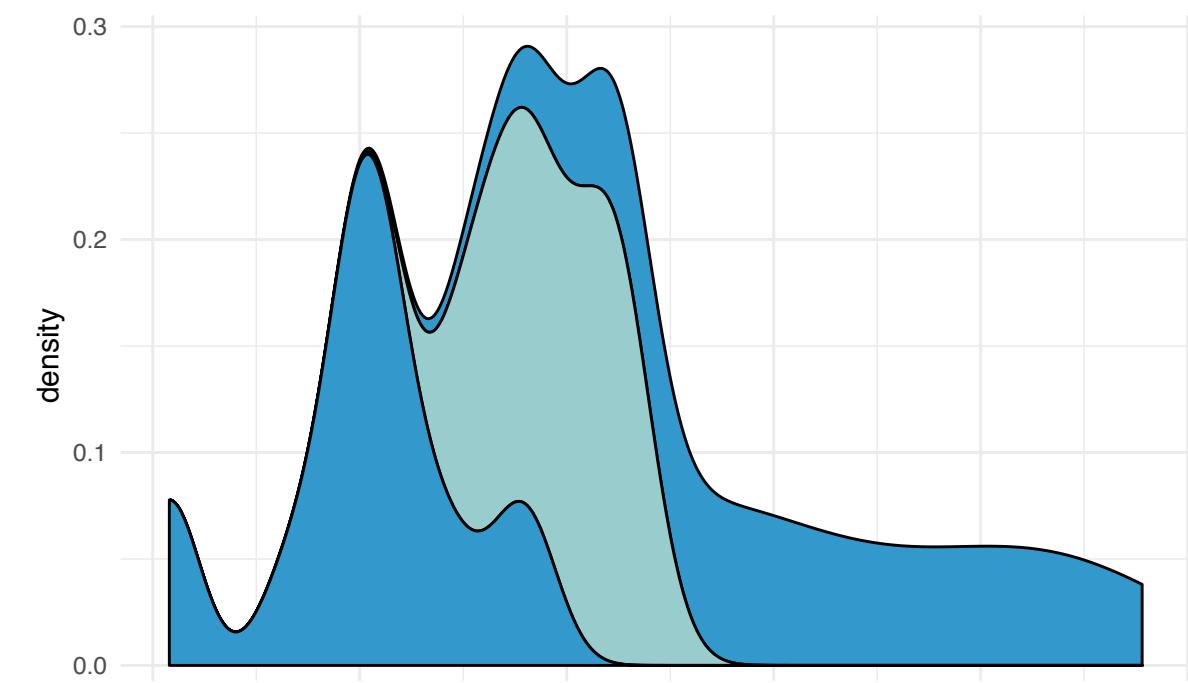
```
ggplot(mtcars) +  
  geom_density(aes(  
    x = mpg,  
    fill = cyl),  
  position = "stack")
```



# Problem 1: vis shows incorrect probability distribution



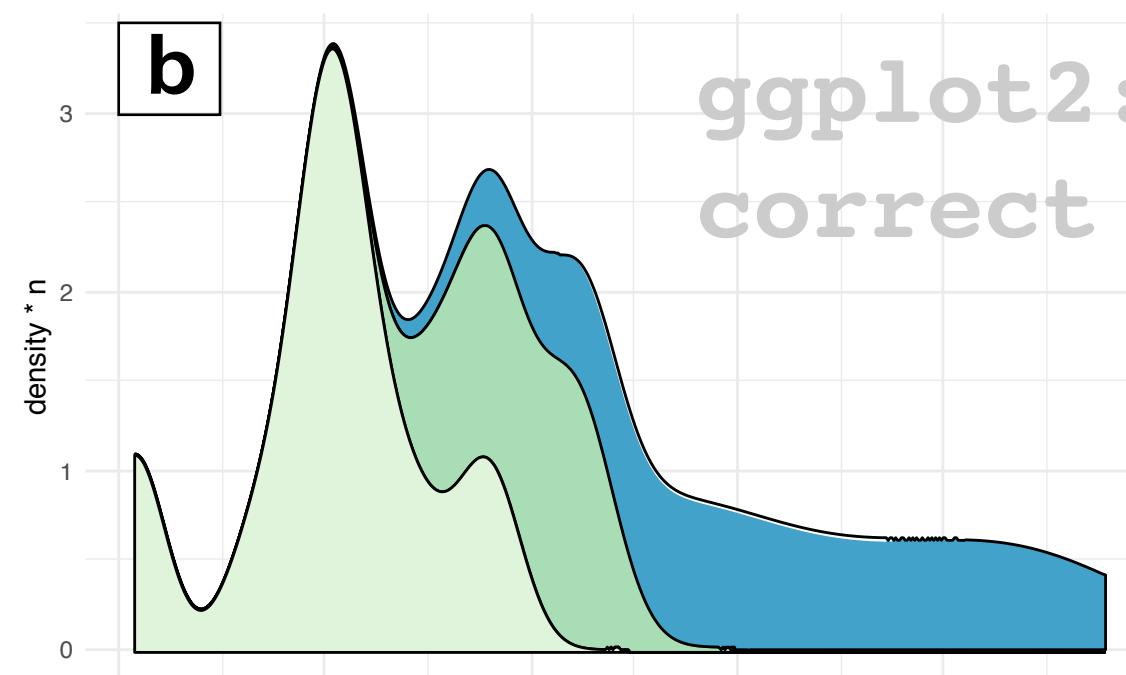
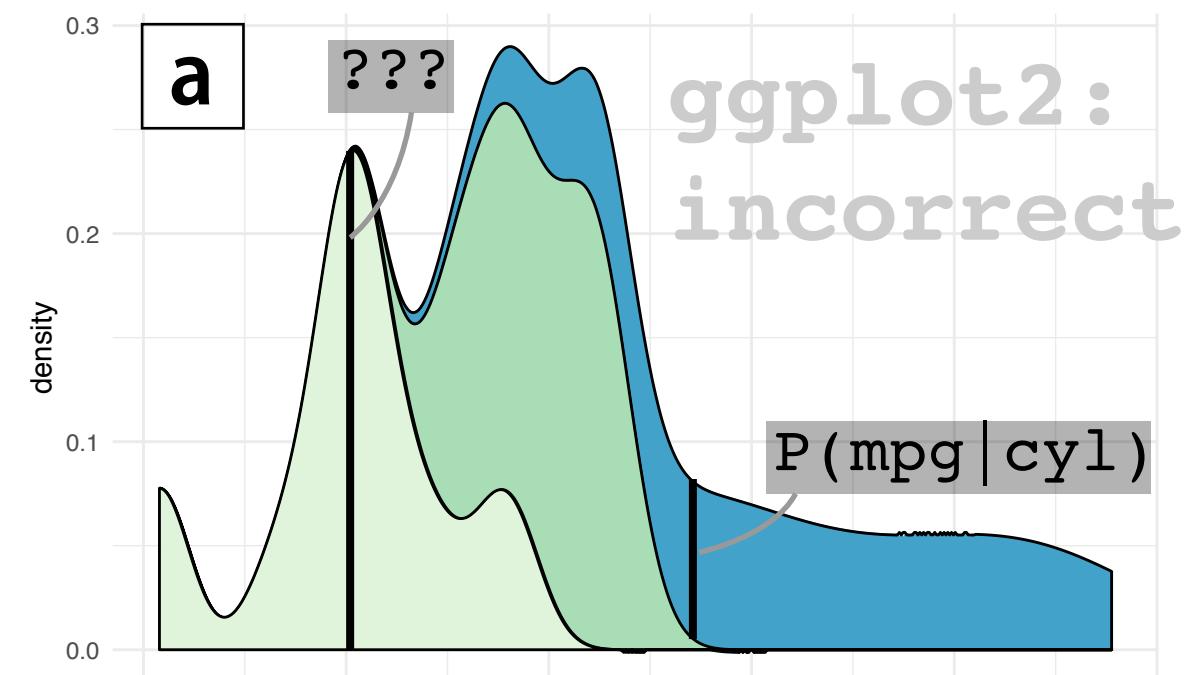
Stack 'em  
→



```
ggplot(mtcars) +  
  geom_density(aes(  
    x = mpg,  
    fill = cyl),
```

```
ggplot(mtcars) +  
  geom_density(aes(  
    x = mpg,  
    fill = cyl),  
  position = "stack")
```

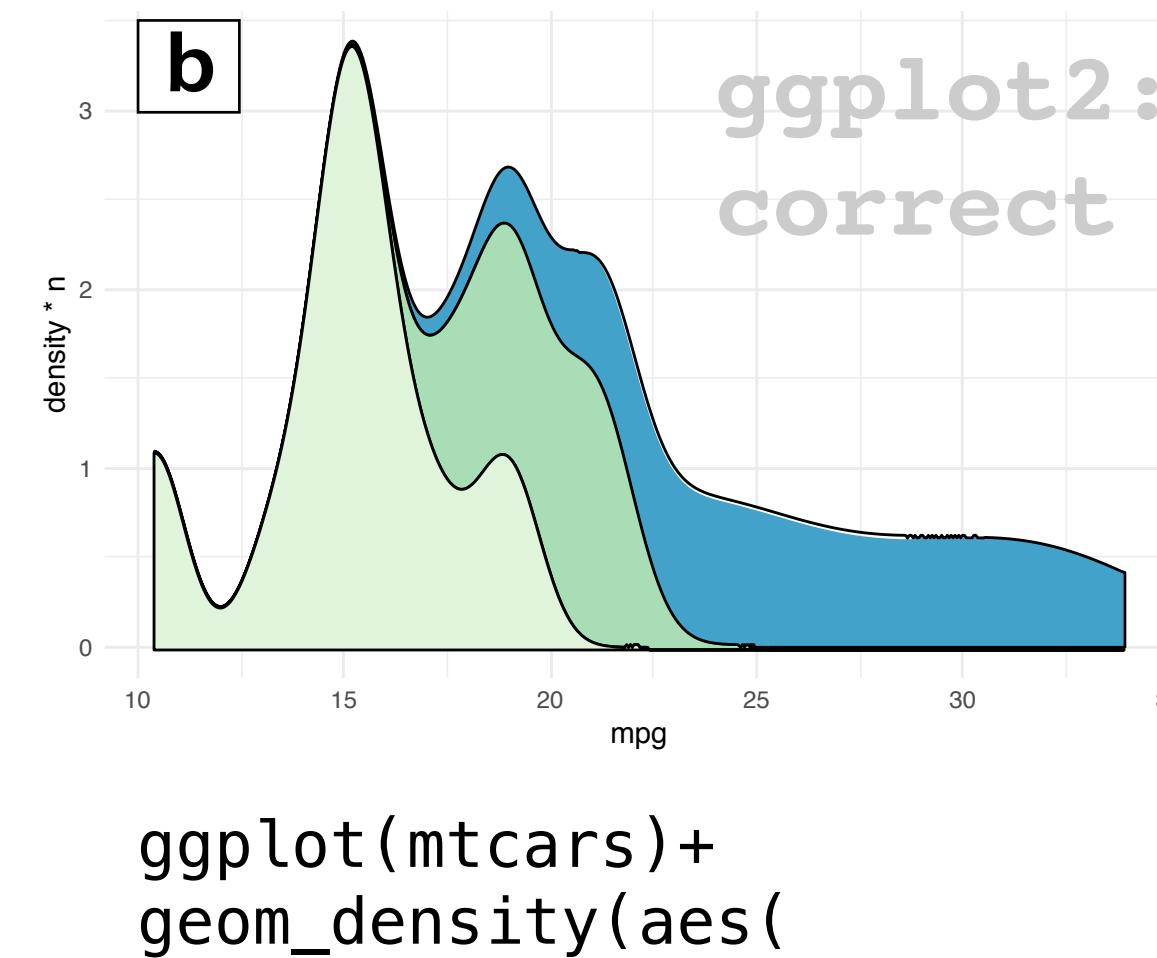
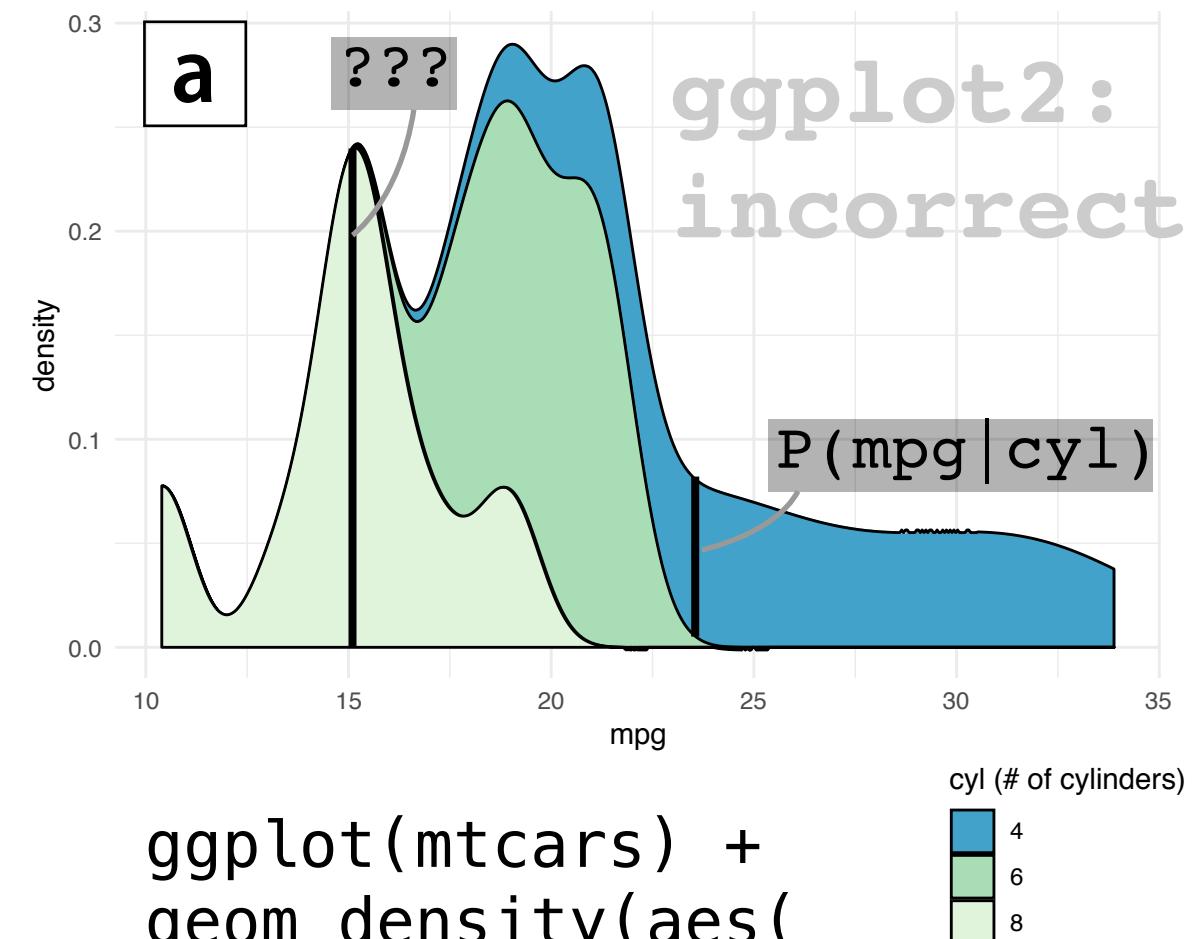
# Problem 1: vis shows incorrect probability distribution



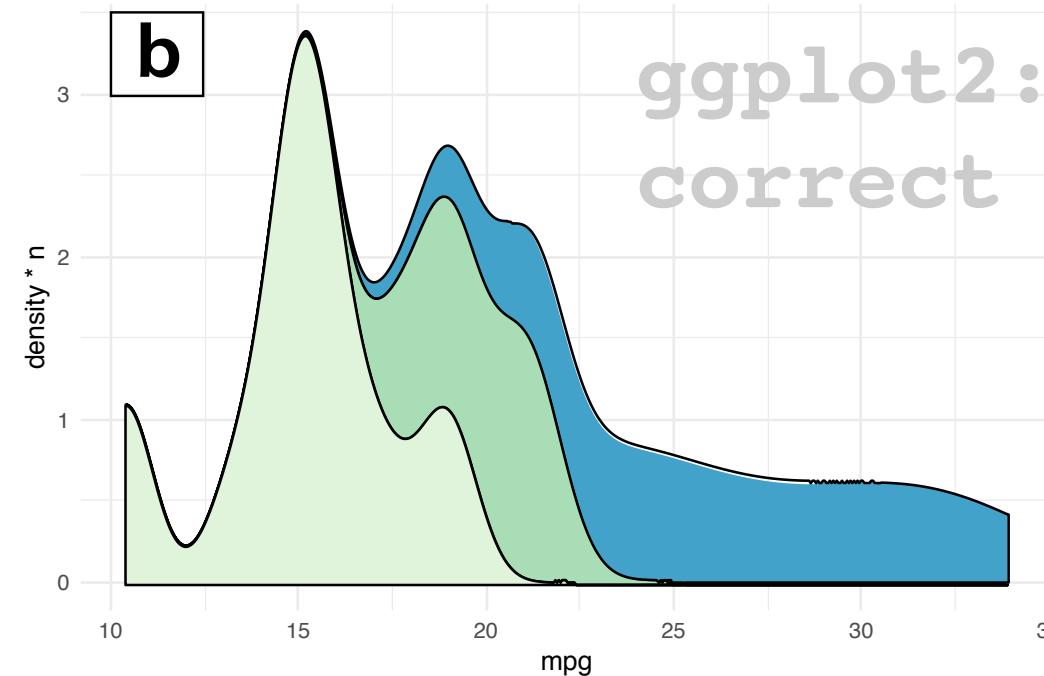
Two problems

1. Wrong distribution of # of cylinders  $P(\text{cyl})$
2. Wrong overall distribution of mileage  $P(\text{mpg})$

# Wait we can fix this density plot



# Problem 2: specifying probability distributions is convoluted

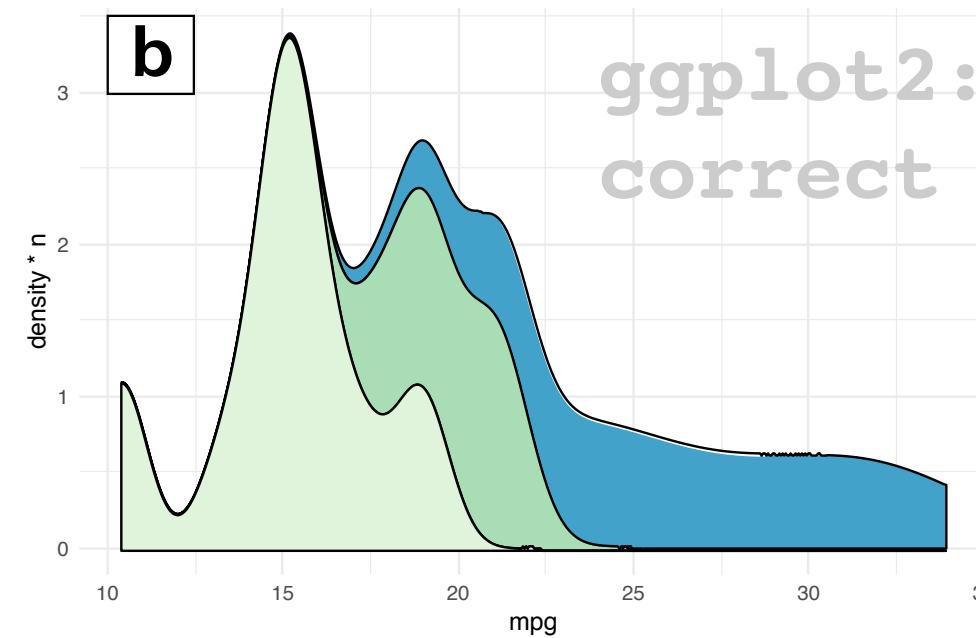


```
ggplot(mtcars)+  
  geom_density(aes(  
    x = mpg,  
    y = stat(density*n),  
    fill = cyl)) +  
  position = "stack")
```

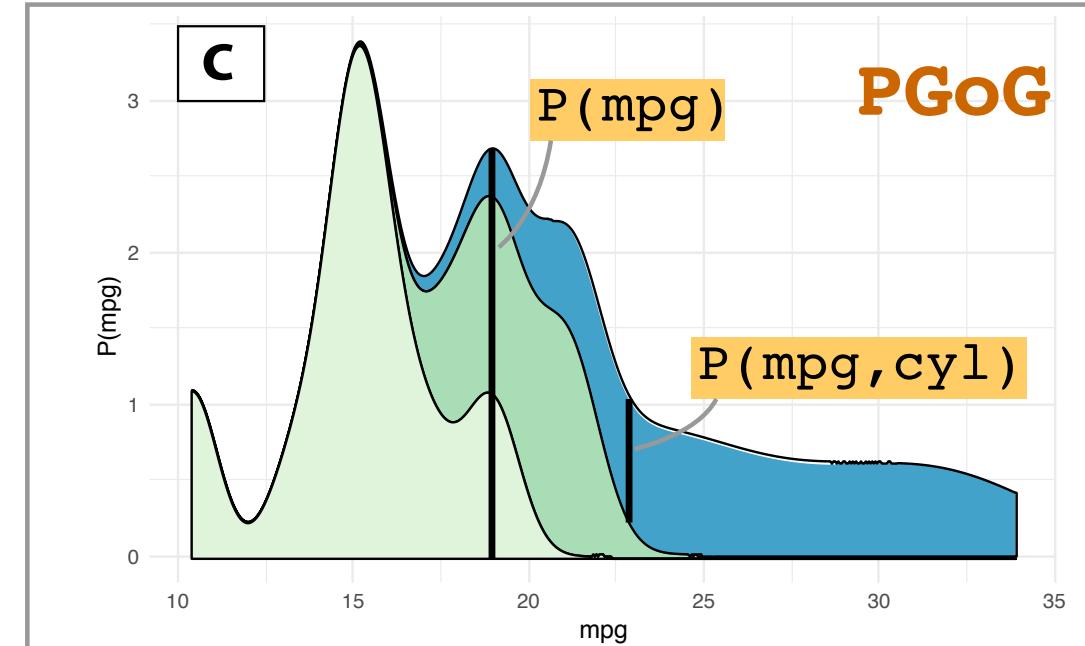
`stat(density * n)`  
 $P(\text{mpg} | \text{cyl} = 4) P(\text{cyl} = 4)$   
 $+ P(\text{mpg} | \text{cyl} = 6) P(\text{cyl} = 6)$   
 $+ P(\text{mpg} | \text{cyl} = 8) P(\text{cyl} = 8)$   
 $= \text{Sum}(P(\text{mpg}, \text{cyl}))$   
 $= P(\text{mpg})$

But what are `stat(density*n)` and `position`?

# Problem 2 can be solved with ...



```
ggplot(mtcars)+  
  geom_density(aes(  
    x = mpg,  
    y = stat(density*n),  
    fill = cyl)) +  
  position = "stack")
```



```
ggplot(mtcars) +  
  geom_bloc(aes(  
    x = mpg,  
    height = P(mpg | cyl) P(cyl),  
    fill = cyl))
```

Details later

# PGoG

Given

1. The need to visualize *probability distributions*
2. Specifying probability distributions is convoluted and error-prone

## A Probabilistic Grammar of Graphics

- A visualization grammar that makes probability distributions first-class citizens
- Unifies a meaningful set of probabilistic visualizations
- Cognitively ergonomic and guaranteed to be correct

# Outline

PGoG in context of

- visualization specification grammar/languages
- formats for communicating probability distributions

Design Requirements  
for PGoG

PGoG abstract grammar

PGoG implementation

Evaluation in terms of

- Expressiveness
- Generativeness
- Cognitive ergonomics

Future: quantitative uncertainty communication!

# Related: how to specify a visualization (Grammar of Graphics)

Data +

example\_df

---

| A | B | C |
|---|---|---|
| 1 | 2 | a |
| 2 | 1 | a |
| 3 | 4 | b |
| 4 | 2 | b |

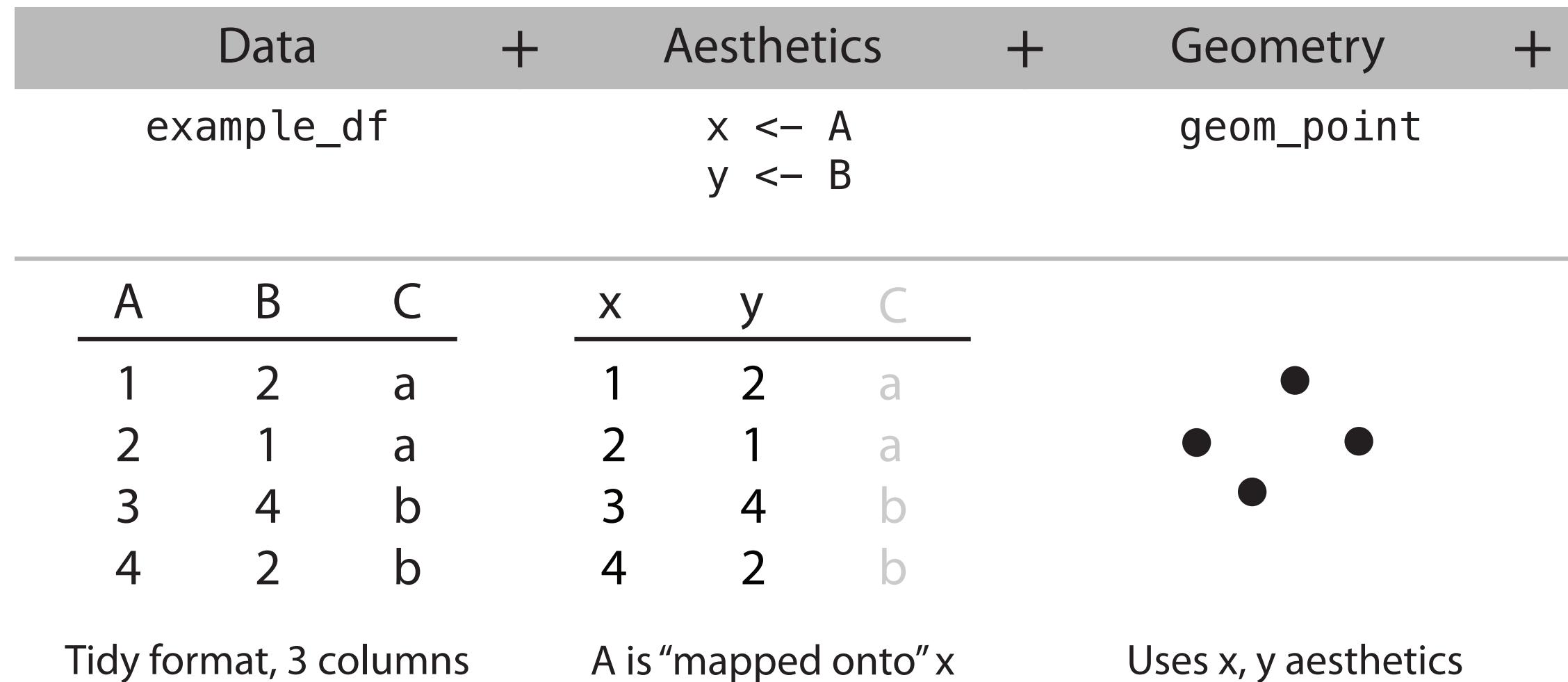
Tidy format, 3 columns

# Related: how to specify a visualization (Grammar of Graphics)

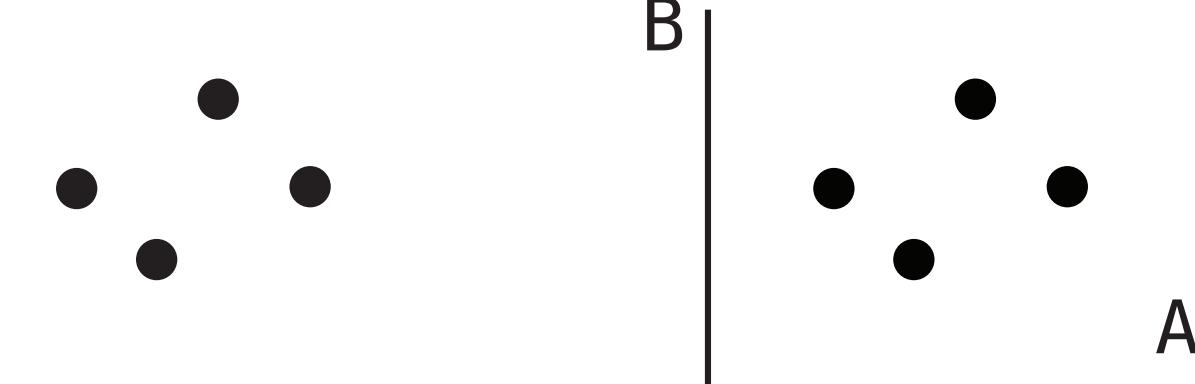
| Data       | + | Aesthetics       | + |
|------------|---|------------------|---|
| example_df |   | x <- A<br>y <- B |   |
| <hr/>      |   |                  |   |
| A          | B | C                | x |
| 1          | 2 | a                | 1 |
| 2          | 1 | a                | 2 |
| 3          | 4 | b                | 3 |
| 4          | 2 | b                | 4 |
|            |   |                  | y |
|            |   |                  | 2 |
|            |   |                  | 1 |
|            |   |                  | 4 |
|            |   |                  | 2 |
|            |   | C                |   |
|            |   | a                |   |
|            |   | a                |   |
|            |   | b                |   |
|            |   | b                |   |

Tidy format, 3 columns      A is “mapped onto” x

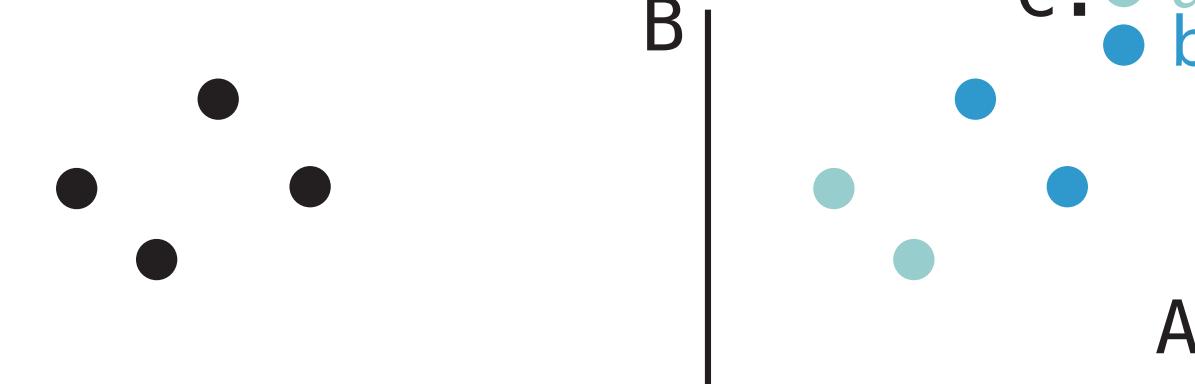
# Related: how to specify a visualization (Grammar of Graphics)

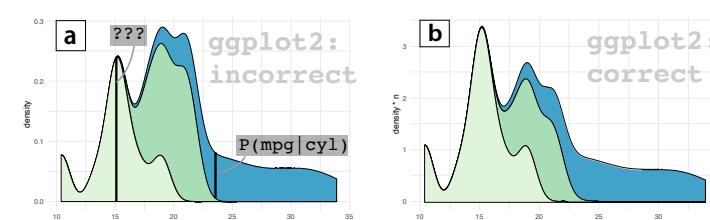


# Related: how to specify a visualization (Grammar of Graphics)

| Data                                                                                                                                                                                                                                             | +                    | Aesthetics           | +              | Geometry   | + ... = A plot |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|----------------------|----------------|------------|----------------|---|---|---|---|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------------------------------------------------------------------------------------|-------|--|
| example_df                                                                                                                                                                                                                                       |                      | x <- A<br>y <- B     |                | geom_point |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| <hr/>                                                                                                                                                                                                                                            |                      |                      |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| <table><thead><tr><th>A</th><th>B</th><th>C</th></tr></thead><tbody><tr><td>1</td><td>2</td><td>a</td></tr><tr><td>2</td><td>1</td><td>a</td></tr><tr><td>3</td><td>4</td><td>b</td></tr><tr><td>4</td><td>2</td><td>b</td></tr></tbody></table> | A                    | B                    | C              | 1          | 2              | a | 2 | 1 | a | 3 | 4 | b | 4 | 2 | b | <table><thead><tr><th>x</th><th>y</th><th>C</th></tr></thead><tbody><tr><td>1</td><td>2</td><td>a</td></tr><tr><td>2</td><td>1</td><td>a</td></tr><tr><td>3</td><td>4</td><td>b</td></tr><tr><td>4</td><td>2</td><td>b</td></tr></tbody></table> | x | y | C | 1 | 2 | a | 2 | 1 | a | 3 | 4 | b | 4 | 2 | b |  | <hr/> |  |
| A                                                                                                                                                                                                                                                | B                    | C                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 1                                                                                                                                                                                                                                                | 2                    | a                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 2                                                                                                                                                                                                                                                | 1                    | a                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 3                                                                                                                                                                                                                                                | 4                    | b                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 4                                                                                                                                                                                                                                                | 2                    | b                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| x                                                                                                                                                                                                                                                | y                    | C                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 1                                                                                                                                                                                                                                                | 2                    | a                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 2                                                                                                                                                                                                                                                | 1                    | a                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 3                                                                                                                                                                                                                                                | 4                    | b                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| 4                                                                                                                                                                                                                                                | 2                    | b                    |                |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |
| Tidy format, 3 columns                                                                                                                                                                                                                           | A is "mapped onto" x | Uses x, y aesthetics | A scatter plot |            |                |   |   |   |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |                                                                                      |       |  |

# Related: how to specify a visualization (Grammar of Graphics)

| Data                                                                                                                                                                                                                                             | + | Aesthetics                     | + | Geometry             | + ... = A plot |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|--------------------------------|---|----------------------|----------------|---|---|---|---|---|---|---|---|---|---|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|-------|---|---|---|---|---|---|---|---|---|---|---|---|--|--------------------------------------------------------------------------------------|--|
| example_df                                                                                                                                                                                                                                       |   | x <- A<br>y <- B<br>color <- C |   | geom_point           |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| <table><thead><tr><th>A</th><th>B</th><th>C</th></tr></thead><tbody><tr><td>1</td><td>2</td><td>a</td></tr><tr><td>2</td><td>1</td><td>a</td></tr><tr><td>3</td><td>4</td><td>b</td></tr><tr><td>4</td><td>2</td><td>b</td></tr></tbody></table> | A | B                              | C | 1                    | 2              | a | 2 | 1 | a | 3 | 4 | b | 4 | 2 | b |  | <table><thead><tr><th>x</th><th>y</th><th>color</th></tr></thead><tbody><tr><td>1</td><td>2</td><td>a</td></tr><tr><td>2</td><td>1</td><td>a</td></tr><tr><td>3</td><td>4</td><td>b</td></tr><tr><td>4</td><td>2</td><td>b</td></tr></tbody></table> | x | y | color | 1 | 2 | a | 2 | 1 | a | 3 | 4 | b | 4 | 2 | b |  |  |  |
| A                                                                                                                                                                                                                                                | B | C                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 1                                                                                                                                                                                                                                                | 2 | a                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 2                                                                                                                                                                                                                                                | 1 | a                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 3                                                                                                                                                                                                                                                | 4 | b                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 4                                                                                                                                                                                                                                                | 2 | b                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| x                                                                                                                                                                                                                                                | y | color                          |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 1                                                                                                                                                                                                                                                | 2 | a                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 2                                                                                                                                                                                                                                                | 1 | a                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 3                                                                                                                                                                                                                                                | 4 | b                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| 4                                                                                                                                                                                                                                                | 2 | b                              |   |                      |                |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |
| Tidy format, 3 columns                                                                                                                                                                                                                           |   | A is "mapped onto" x           |   | Uses x, y aesthetics | A scatter plot |   |   |   |   |   |   |   |   |   |   |  |                                                                                                                                                                                                                                                      |   |   |       |   |   |   |   |   |   |   |   |   |   |   |   |  |                                                                                      |  |



# Related: how to specify a visualization (layout-based is viscous)

(Blackwell et al. 2001)

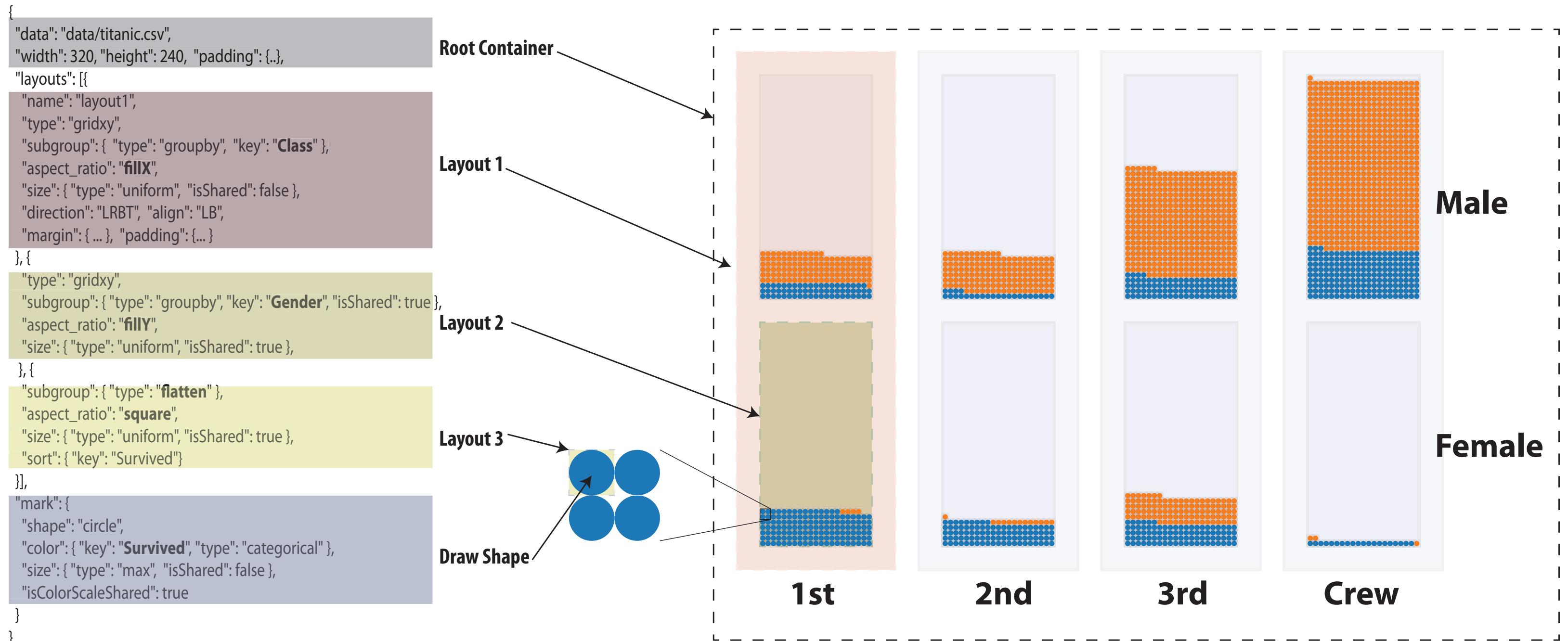
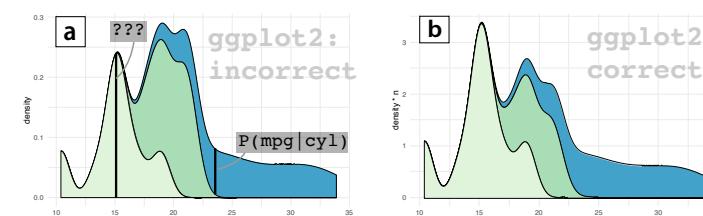
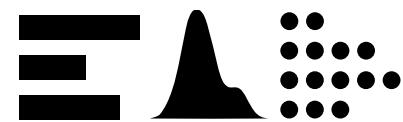


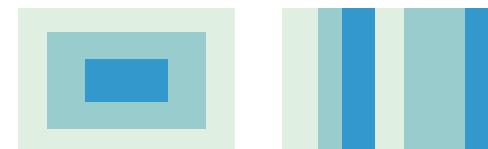
Fig. 6. Example grammar to generate a unit column chart for survivors of the Titanic by passenger class. (Park et al. 2017)

# Related: how to specify a visualization in general

## Grammar of Graphics



## Layout-based



Correct?  
Ergonomic?

## Constraints-based

```
encoding(e1).  
:- not channel(e1,x).  
:- not field(e1,horsepower).  
:- not bin(e1,_).
```

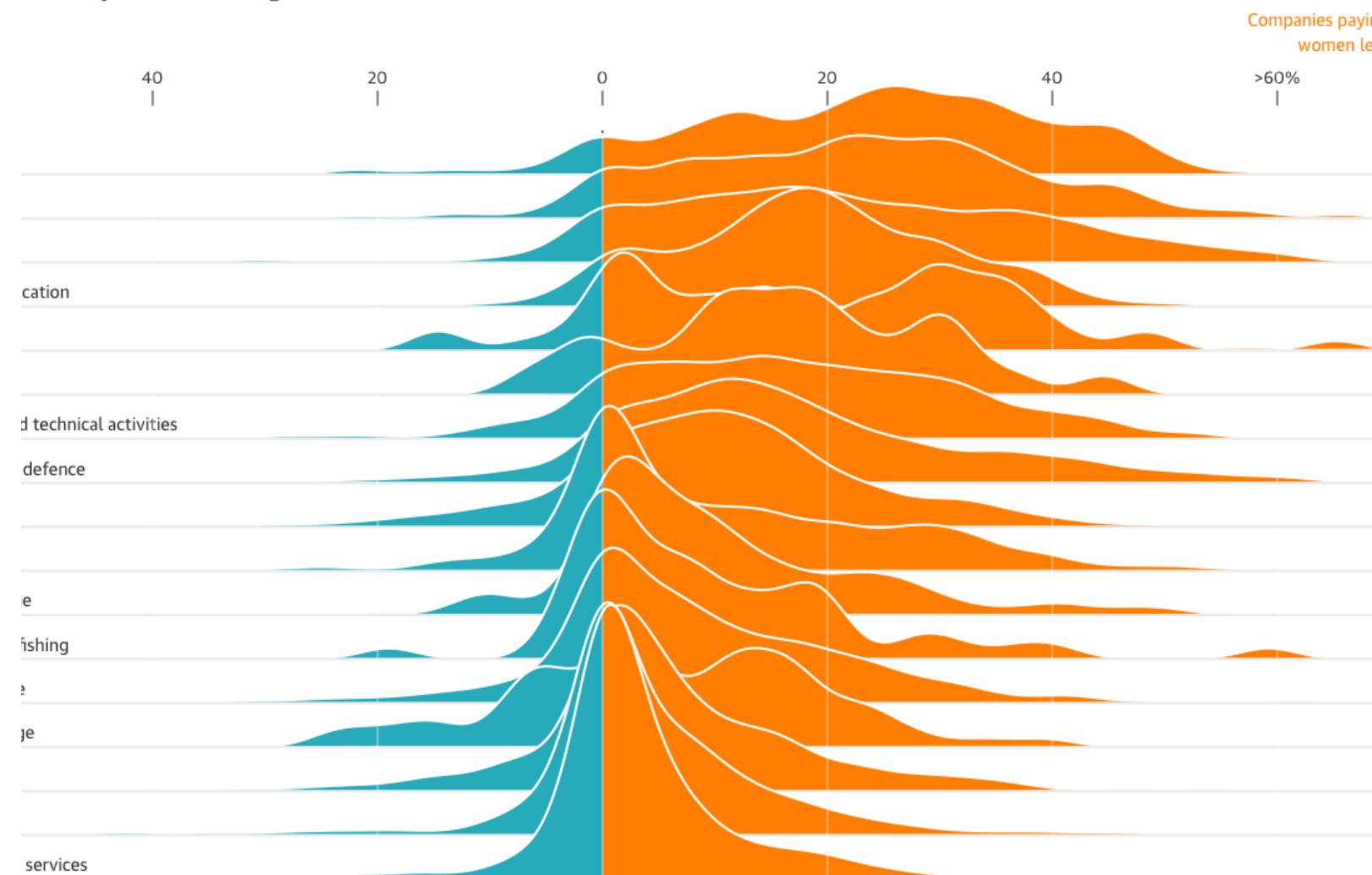
(Moritz et al. 2019)

Need a closer integration  
between statistics and  
visualization (Heer and Shneiderman 2012)

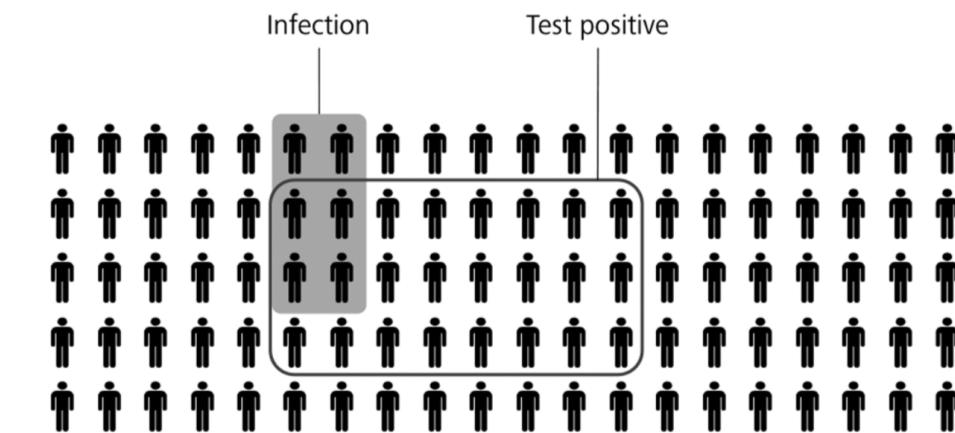
# Related: communicating/visualizing uncertainty

Probabilistic visualizations  
are often used to communicate uncertainty data

**Women are likely to be underpaid in certain sectors**



<https://www.theguardian.com/news/ng-interactive/2018/apr/05/women-are-paid-less-than-men-heres-how-to-fix-it>

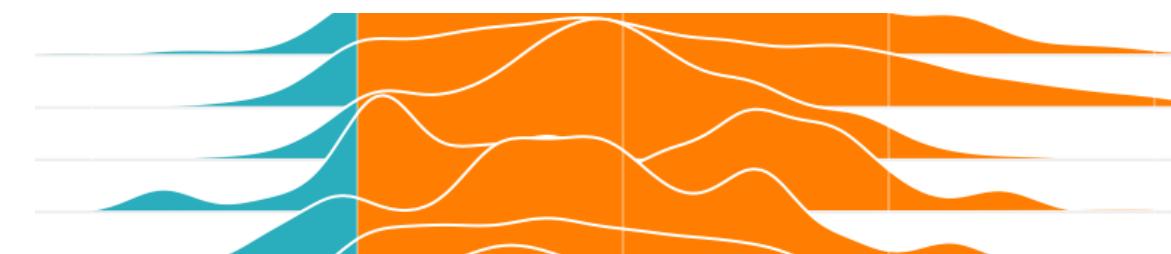
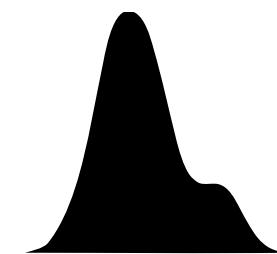


# Related: communicating/visualizing uncertainty

## Probabilistic visualizations

are often used to communicate uncertainty data

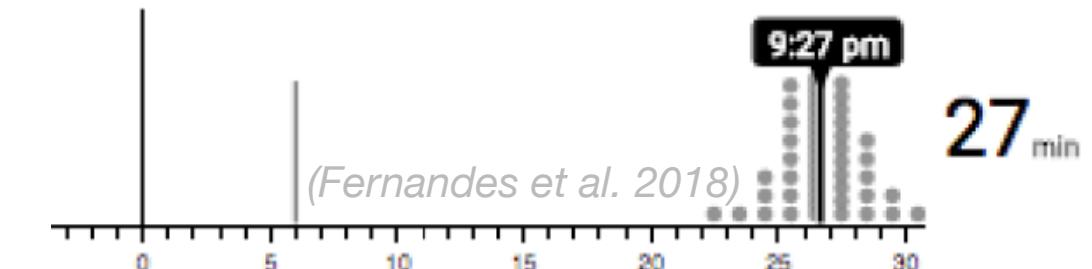
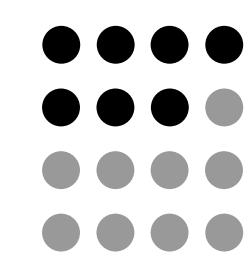
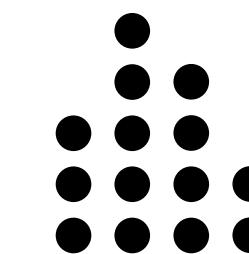
*Probability format    X%*



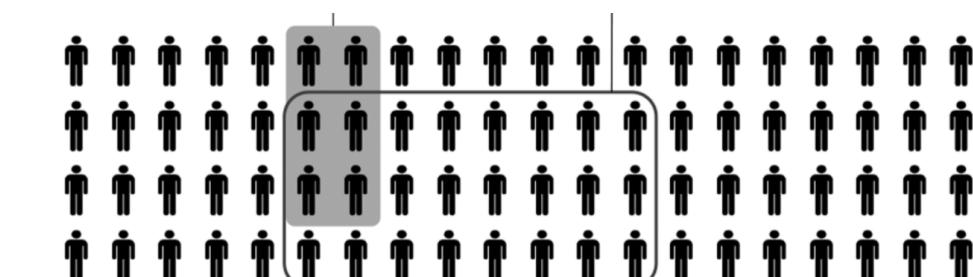
(Gigerenzer and Hoffrage 1995)

(Guo et al. 2019)

*Frequency format    X-in-100*



(Fernandes et al. 2018)

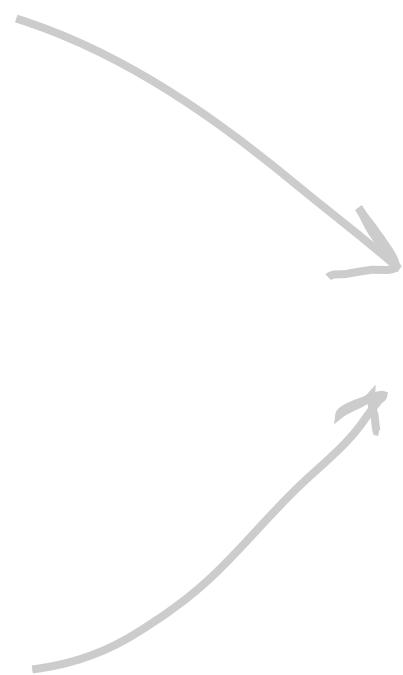


(Binder, Krauss, and Bruckmaier 2015)

*Which one to choose?*

A closer integration  
between statistics and  
visualization

A need to support  
probability and  
frequency formats



## Design Requirements for a probabilistic Grammar of Graphics

A closer integration  
between statistics and  
visualization

- 
- 1 Guaranteeing correctness  
of distributions expressed in  
visualization
  - 2 Enabling specification  
cognitively ergonomic  
and close to probability  
expressions, such as  $P(A|B)$

## Design Requirements for a probabilistic Grammar of Graphics

A need to support  
probability and  
frequency formats

3 Facilitating exploration *with  
coherent and reusable grammar  
components*

4 (and automation in the future)

## Design Requirements for a probabilistic Grammar of Graphics



# The design process

Defaults

Data  $\dashrightarrow A$

Aesthetics  $\rightarrow x \leftarrow A$

Layer

Data

Aesthetics

Geom  $\dashrightarrow \text{geom\_bar}$

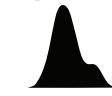
Stat

Position

Scale



geom\_density



geom\_points



geom\_rect



geom\_...

(Wickham 2010)

# What is the Probabilistic Grammar of Graphics?

| Grammar    | ggplot2                                            | PGoG                                     |
|------------|----------------------------------------------------|------------------------------------------|
| Defaults   |                                                    |                                          |
| Data       | $\text{Data} \dashrightarrow A$                    | $P(A B, \dots)$                          |
| Aesthetics | $\text{Aesthetics} \dashrightarrow x \leftarrow A$ | $\text{height} \leftarrow P(A B, \dots)$ |
| Layer      |                                                    |                                          |
| Data       |                                                    |                                          |
| Aesthetics |                                                    |                                          |
| Geom       | $\text{Geom} \dashrightarrow \text{geom\_bar}$     | $\text{geom\_bloc}$                      |
| Stat       |                                                    |                                          |
| Position   |                                                    | $\text{geom\_icon}$                      |
| Scale      |                                                    |                                          |
| Coord      | $\text{geom\_density}$                             |                                          |
| Facet      |                                                    |                                          |
|            | $\text{geom\_points}$                              |                                          |
|            |                                                    |                                          |
|            | $\text{geom\_rect}$                                |                                          |
|            |                                                    |                                          |
|            | $\text{geom\_...}$                                 |                                          |

(Wickham 2010)

# What is the Probabilistic Grammar of Graphics?

1. The PGoG **grammar** is an extension to *Grammar of Graphics*
2. Probability distributions are first class citizens (data) and other grammar components (aesthetics and geometries) are defined around them

# PGoG Grammar/*data*

|                   | mpg  | cyl | am |
|-------------------|------|-----|----|
| Mazda RX4         | 21.0 | 6   | 1  |
| Mazda RX4 Wag     | 21.0 | 6   | 1  |
| Datsun 710        | 22.8 | 4   | 1  |
| Hornet 4 Drive    | 21.4 | 6   | 0  |
| Hornet Sportabout | 18.7 | 8   | 0  |
| Valiant           | 18.1 | 6   | 0  |

Column variable

mpg

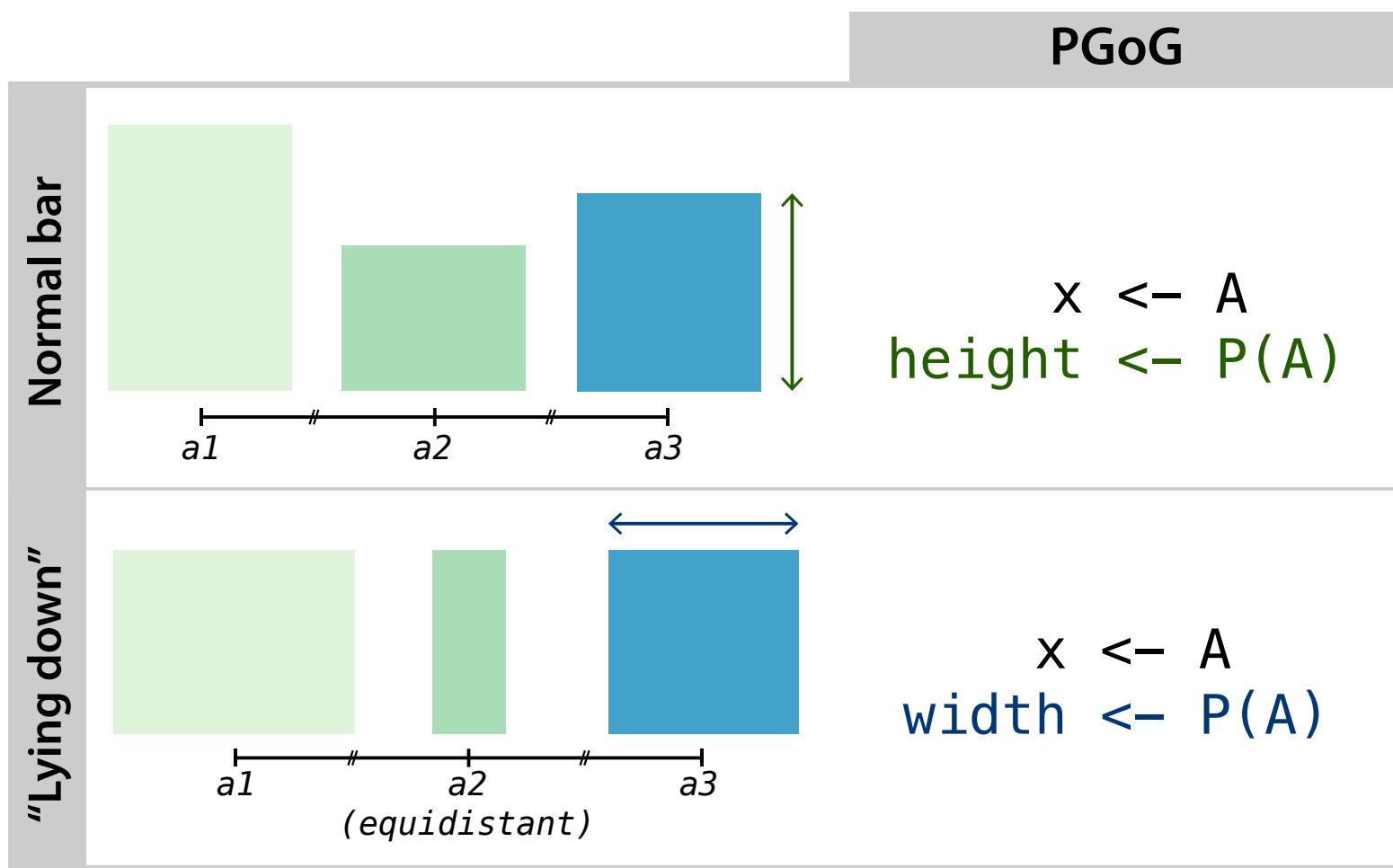
A column in tidy dataset

Probabilistic variable

$P(\text{mpg} | \text{cyl})$

In the form of  $P(A...|B...)$ , where A, B and ... are variables in columns

# PGoG Grammar/aesthetics 1/3

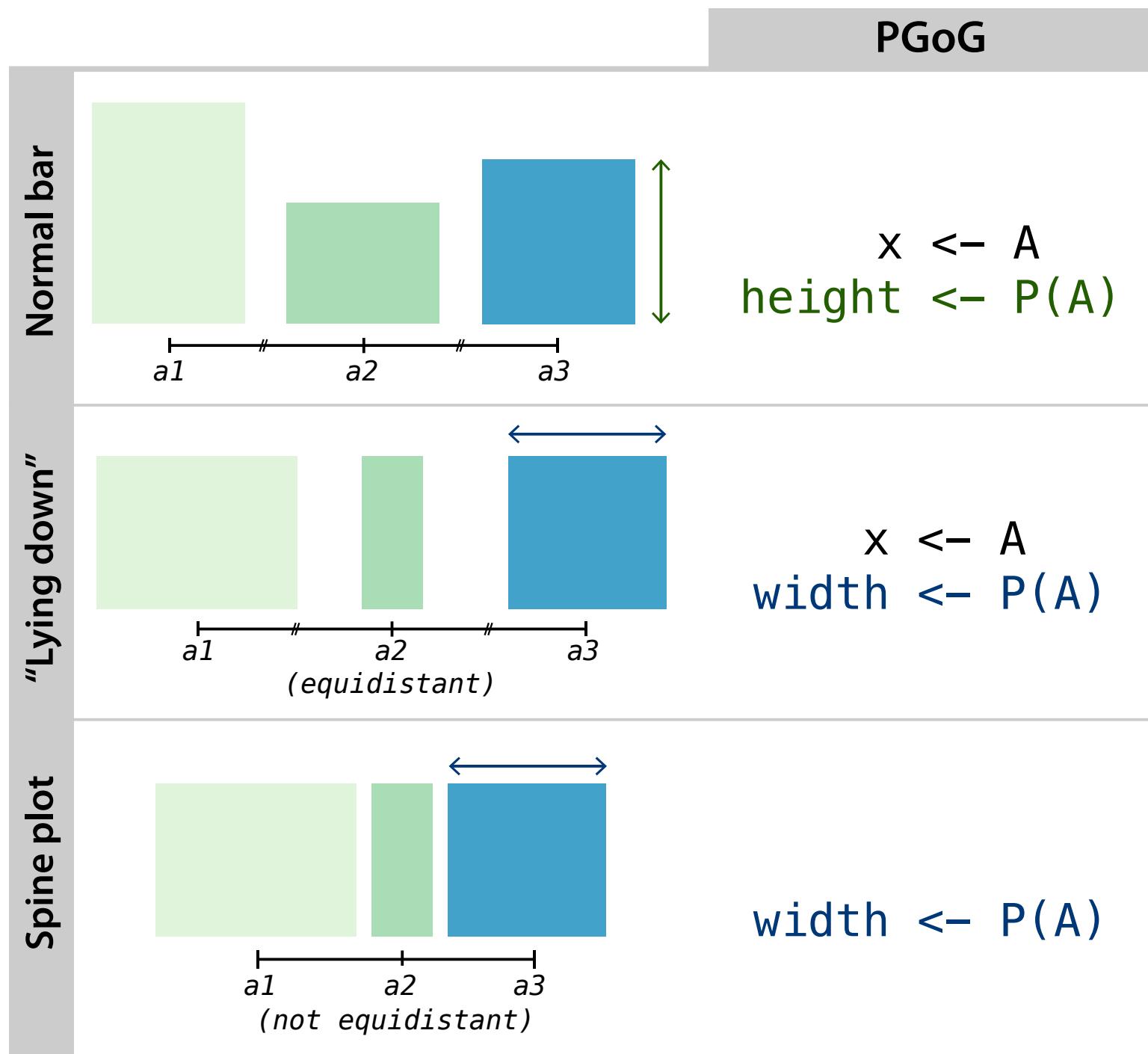


Probabilistic aesthetics

**width, height**

- Works with probabilistic variables only
- Expresses the probability value by length

# PGoG Grammar/aesthetics 2/3



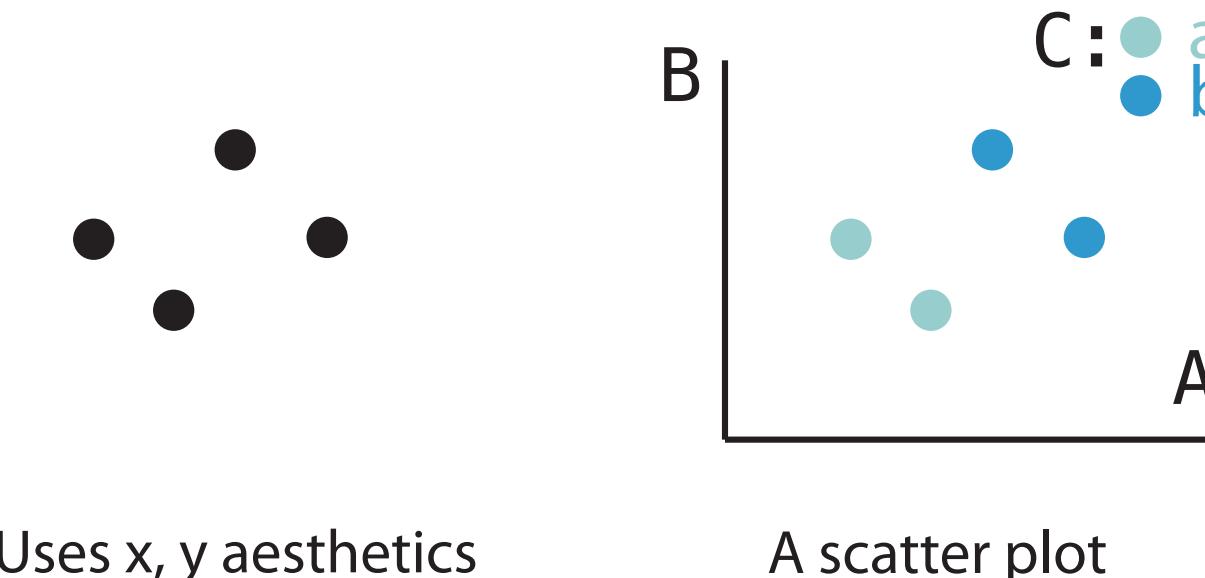
Probabilistic aesthetics

Coordinate aesthetics

**x, y**

- For discrete vars: equidistant partitions
- For continuous vars: as one would expect

# PGoG Grammar/aesthetics 3/3



Uses x, y aesthetics

A scatter plot

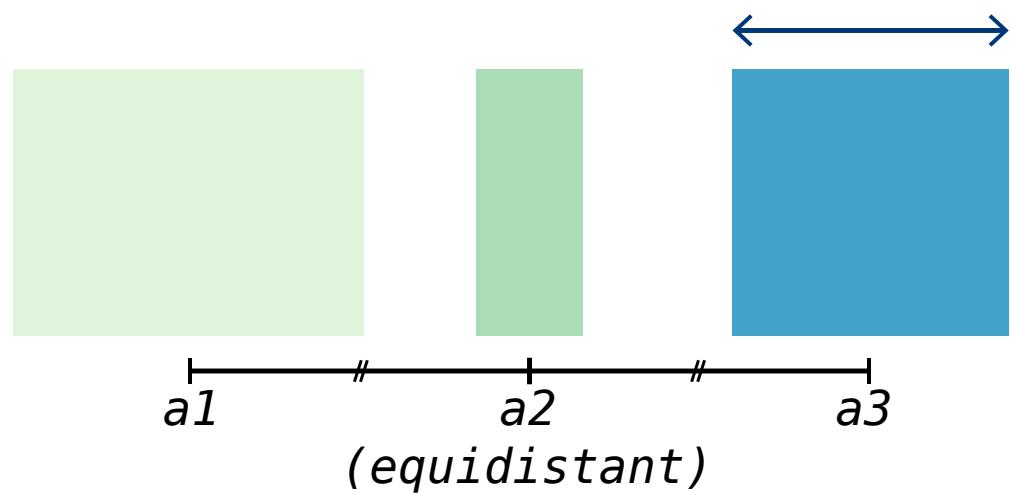
Probabilistic aesthetics

Coordinate aesthetics

Visual aesthetics

fill, color, alpha, ...

# PGoG Grammar/Example for conditional

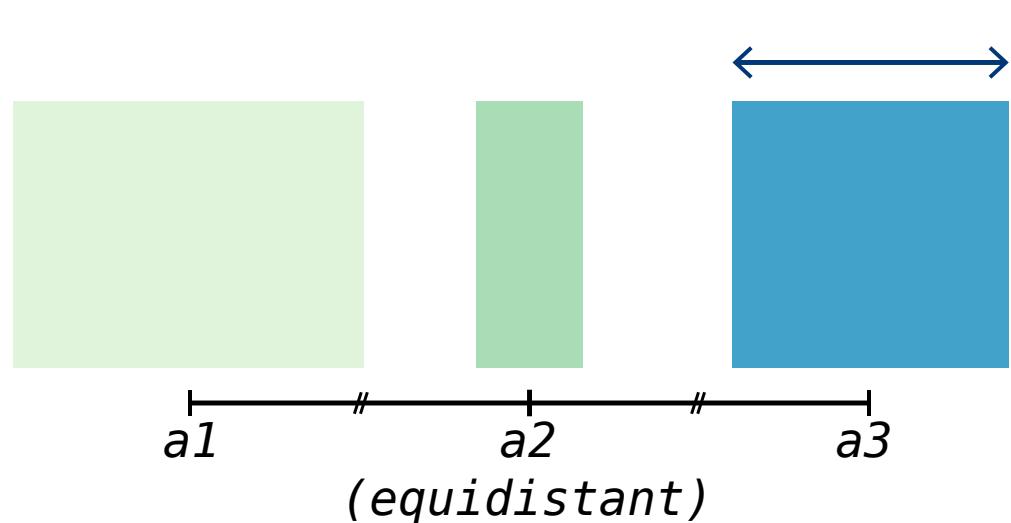


$x \leftarrow A$   
 $\text{width} \leftarrow P(A)$

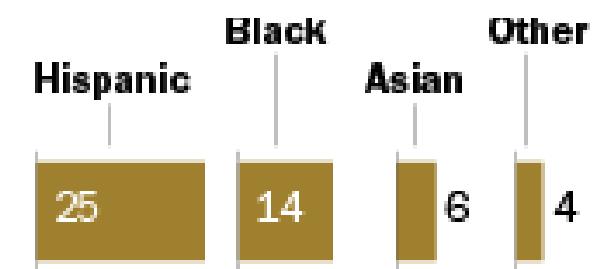
$x \leftarrow \text{race}$

$\text{width} \leftarrow \text{P(race|generation)}$

# PGoG Grammar/Example for conditional



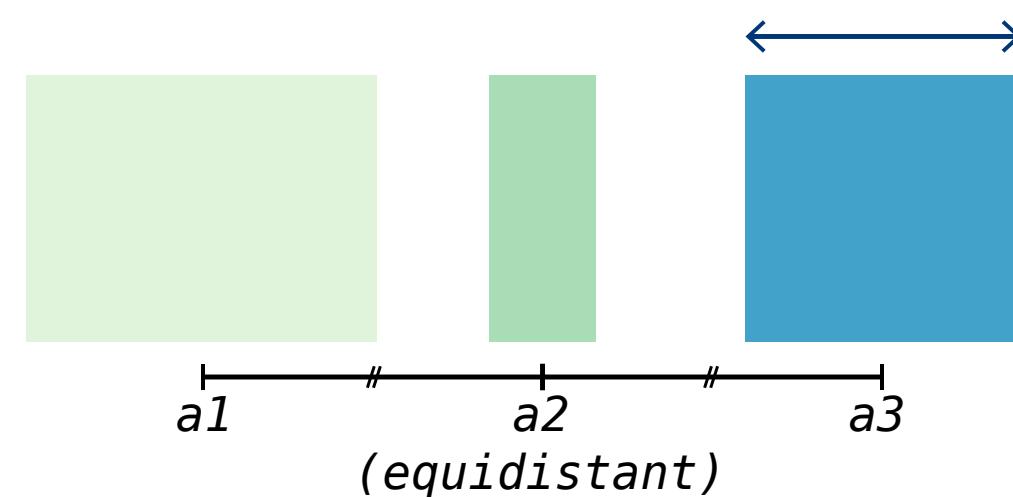
```
x <- A  
width <- P(A)
```



$x <- \text{race}$

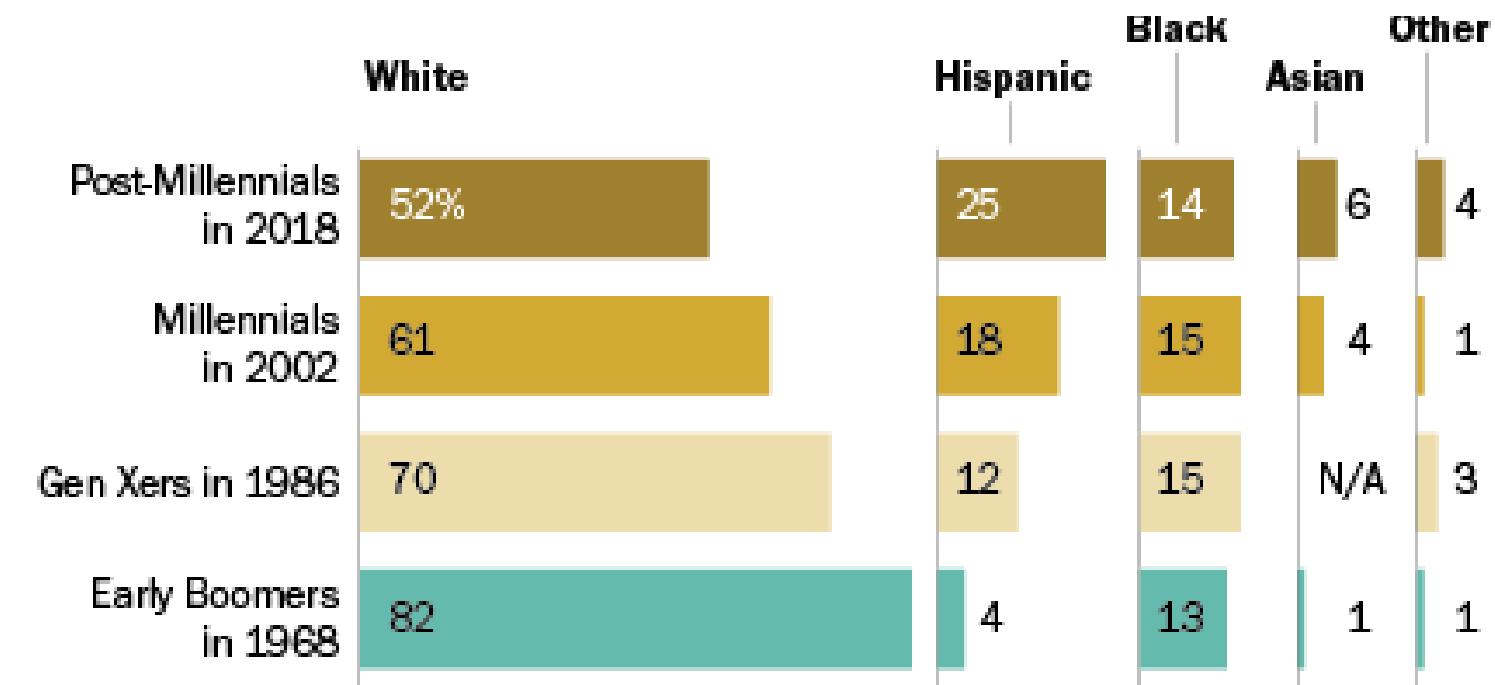
$\text{width} <- \text{P}(\text{race}|\text{generation})$

# PGoG Grammar/Example for conditional



```
x <- A  
width <- P(A)
```

<http://www.pewresearch.org/fact-tank/2018/12/13/18-striking-findings-from-2018/>



```
x <- race  
y <- generation  
width <- P(race|generation)
```

# PGoG Grammar/*Example for joint*

Math

$$P(\text{mpg} | \text{cyl}) P(\text{cyl}) = P(\text{mpg}, \text{cyl})$$

# PGoG Grammar/*Example for joint*

Math

$$P(\text{mpg} | \text{cyl}) \ P(\text{cyl}) = P(\text{mpg}, \text{cyl})$$

Coord aes

`x <- mpg`

# PGoG Grammar/*Example for joint*

Math

$$P(\text{mpg} | \text{cyl}) \ P(\text{cyl}) = P(\text{mpg}, \text{cyl})$$

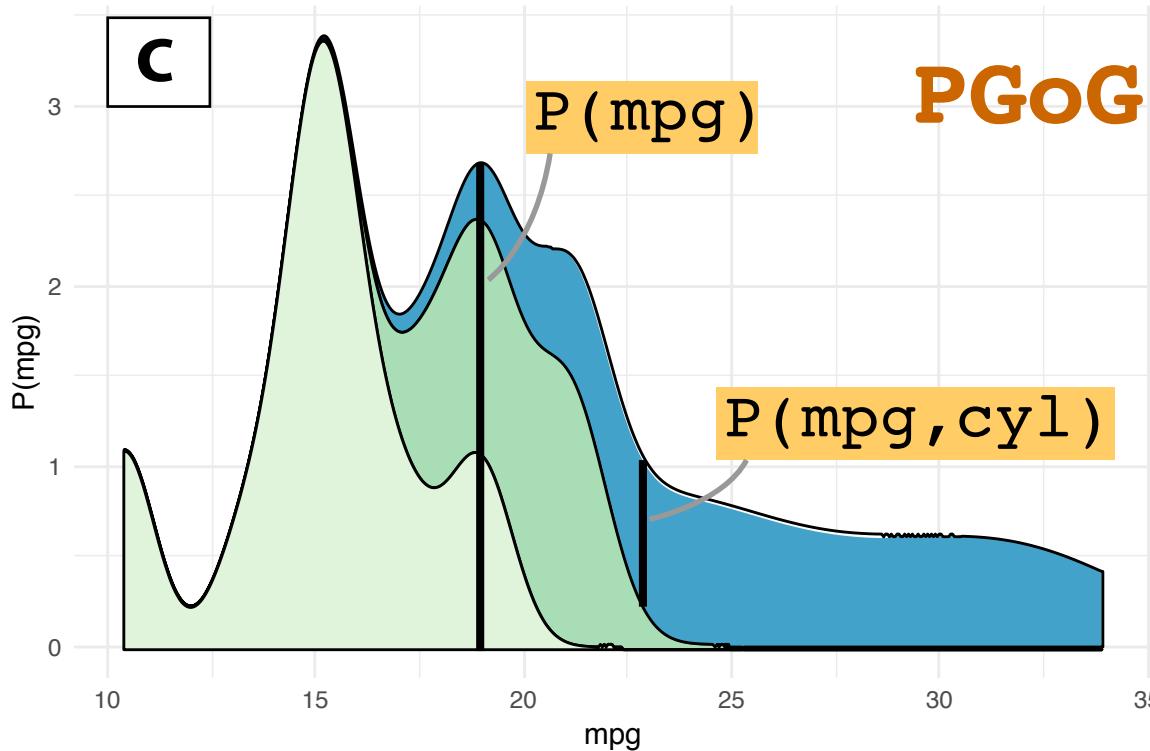
Coord aes

```
x <- mpg
```

Prob aes

```
height <- P(cyl) P(mpg | cyl)
```

# PGoG Grammar/Example for joint



```
ggplot(mtcars) +  
  geom_bloc(aes(  
    x = mpg,  
    height = P(mpg|cyl) P(cyl),  
    fill = cyl))
```

Math

$$P(\text{mpg} | \text{cyl}) P(\text{cyl}) = P(\text{mpg}, \text{cyl})$$

Coord aes

`x <- mpg`

Prob aes

`height <- P(cyl) P(mpg|cyl)`

Visual aes

`fill <- cyl`

# PGoG Grammar/*checking correctness* 1/2

One of the **rules**: the probabilistic variables need to be valid factors of a **probability function**

```
x <- gear  
height <- P(gear|am)  
          P(cyl|gear, am)
```

# PGoG Grammar/*checking correctness* 1/2

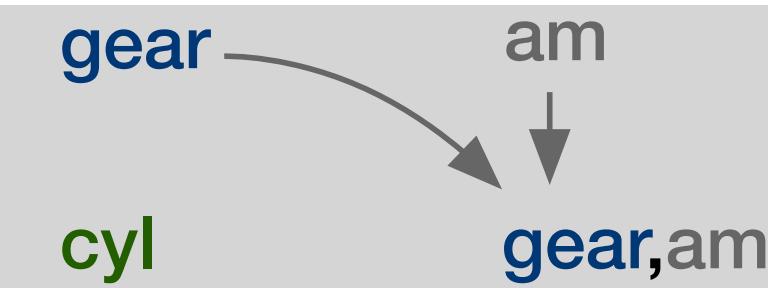
One of the rules: the probabilistic variables need to be valid factors of a **probability function**

$x \leftarrow \text{gear}$   
height  $\leftarrow \begin{cases} P(\text{gear}|\text{am}) \\ P(\text{cyl}|\text{gear}, \text{am}) \end{cases}$

**P(cyl, gear|am)**      **marg**      **cond**

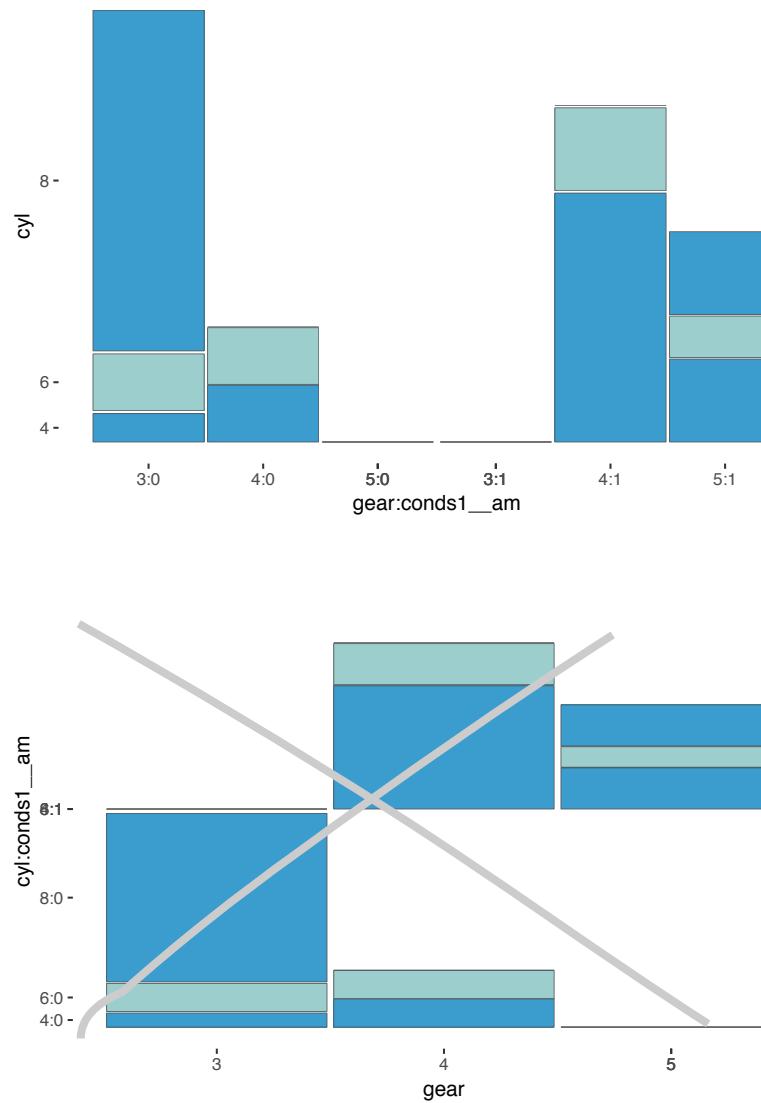
$P(\text{gear}|\text{am})$

$P(\text{cyl}|\text{gear}, \text{am})$



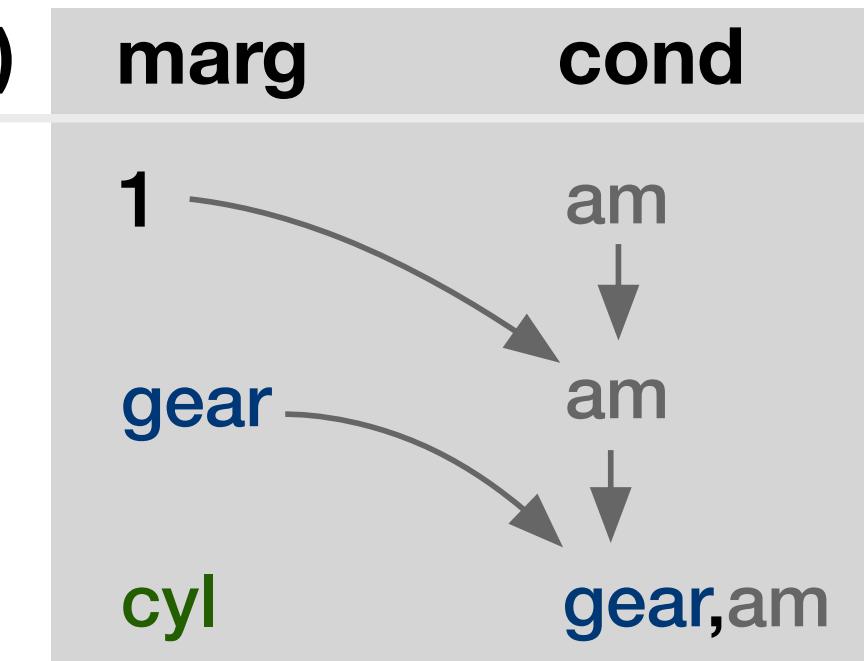
A simple algorithm that sorts and checks probabilistic variables

# PGoG Grammar/checking correctness 2/2



$x \leftarrow \text{gear, am}$   
 $\text{height} \leftarrow P(\text{gear|am})$   
 $P(\text{cyl|gear, am})$

**aes**       $P(\text{cyl, gear|am})$   
 $x \leftarrow P(1|\text{am})$   
 $\text{height} \leftarrow P(\text{gear|am})$   
 $\text{height} \leftarrow P(\text{cyl|gear, am})$



Grammar

ggplot2

PGoG

Defaults

Data  $\dashrightarrow A$

Aesthetics  $\dashrightarrow x \leftarrow A$

Layer

Data

Aesthetics

Geom  $\dashrightarrow \text{geom\_bar}$



Stat  
Position

Scale



Coord

geom\_points



Facet

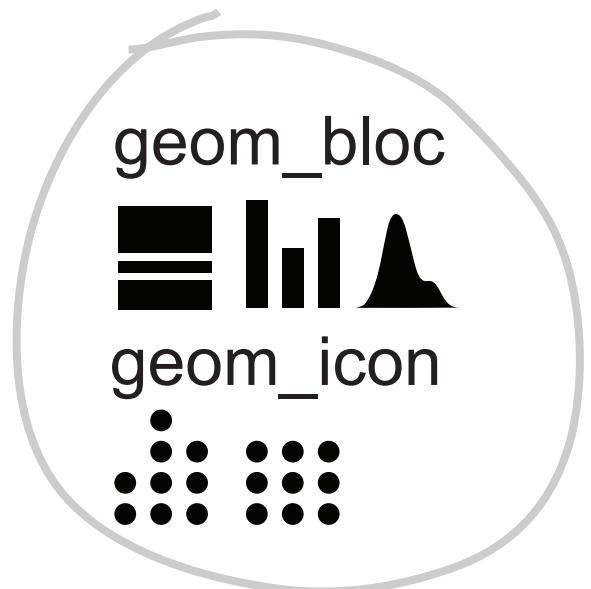
geom\_rect



geom\_...

$P(A|B, \dots)$

height  $\leftarrow P(A|B, \dots)$



## PGoG Grammar/ *geometries* 1/2

---

**ggplot2**

geom\_bar



geom\_mosaic\*



geom\_density



geom\_violin

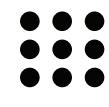


geom\_density\_ridges\*



---

geom\_waffle\*



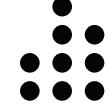
**PGoG**

geom\_bloc

Comparison: look at all those geometries in **ggplot2** we have replaced

**geom\_icon**

geom\_dotplot



---

\* ggplot2 extensions

# PGoG Grammar/*geometries*

`geom_bloc`: recursive  
subpartition to support  
many probabilistic vari-  
ables



`geom_icon` needs a  
new way to pack icons

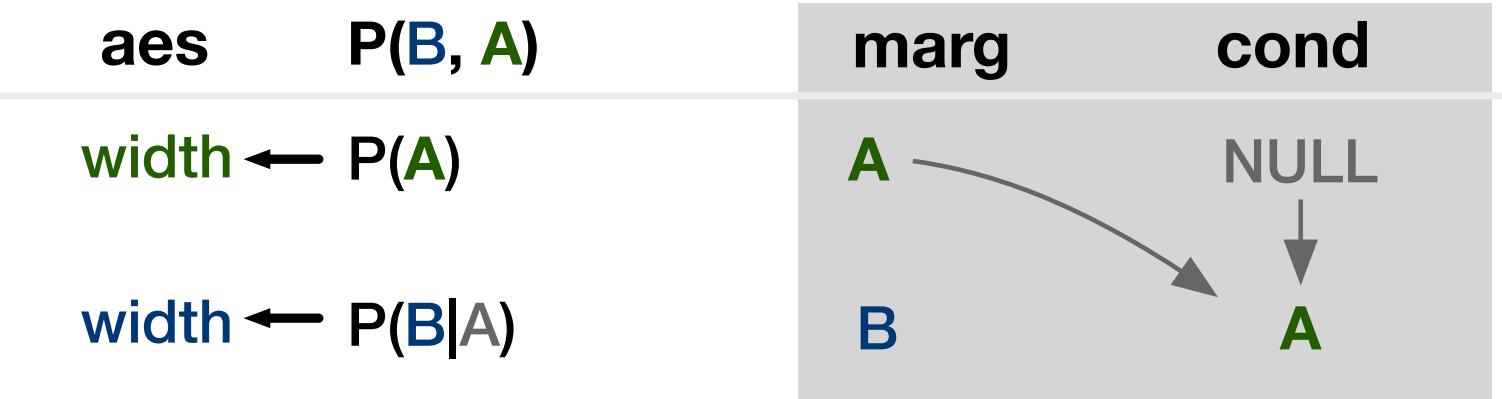
```
y <- A  
width <- P(A) P(B|A)  
fill <- B
```

# PGoG Grammar/*geometries*

`geom_bloc`: recursive  
subpartition to support  
many probabilistic vari-  
ables

`geom_icon` needs a  
new way to pack icons

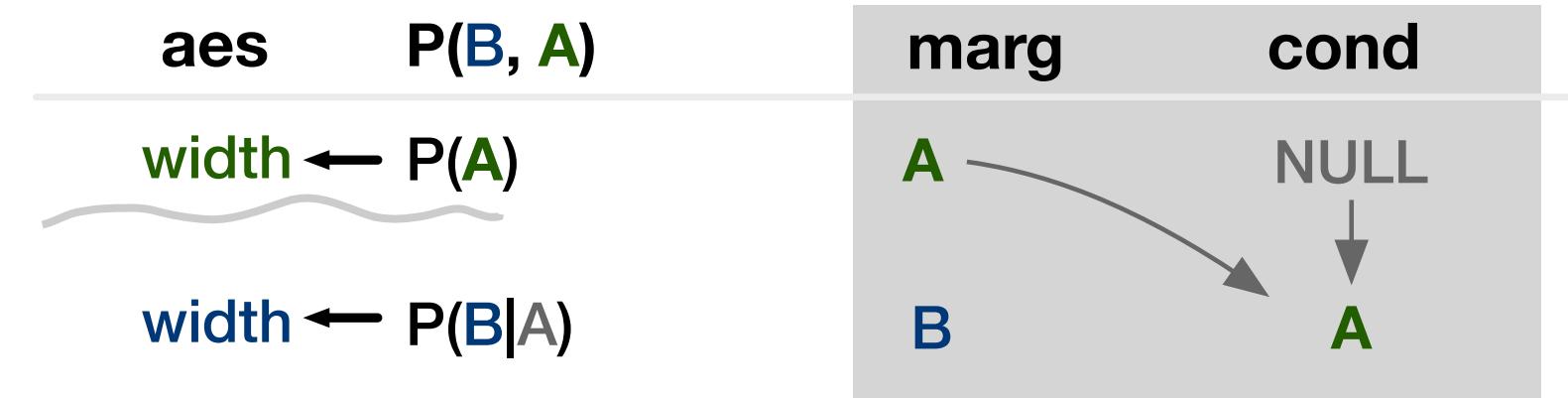
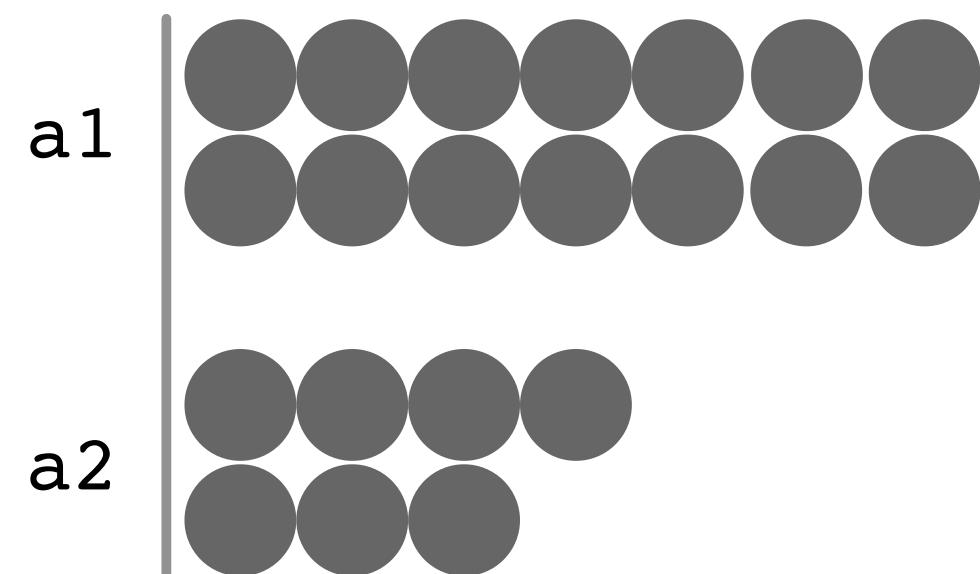
```
y <- A  
width <- P(A) P(B|A)  
fill <- B
```



Probability structure (the “chain”) deter-  
mines the visualization structure

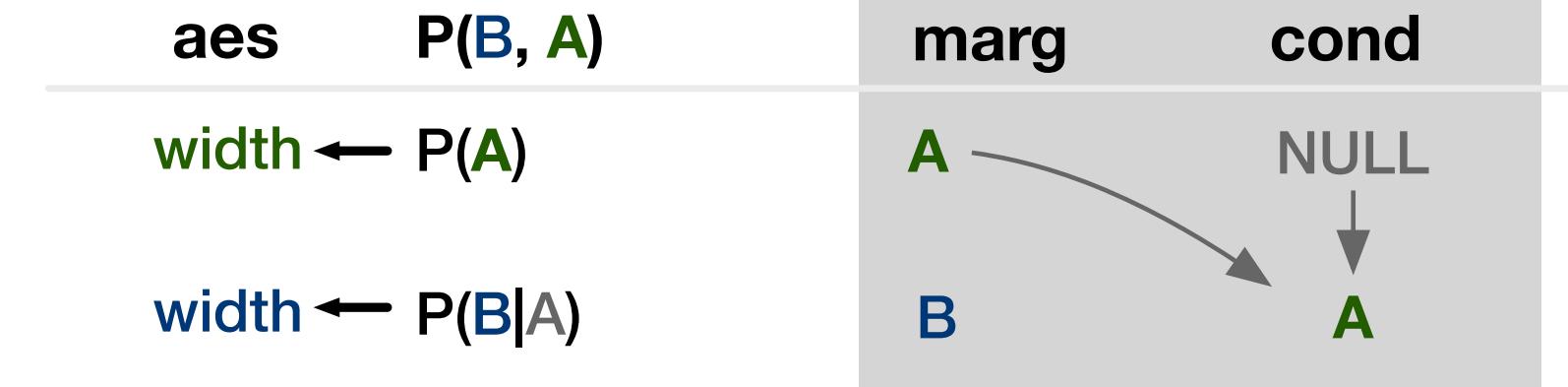
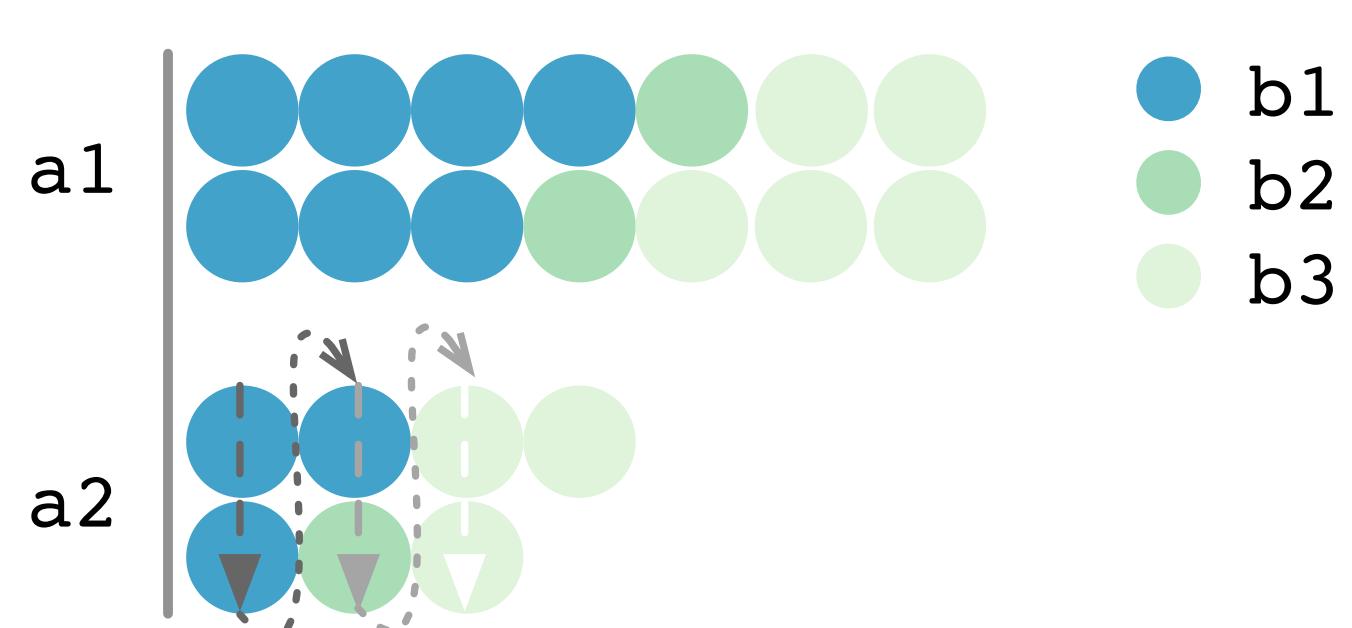
# PGoG Grammar/*geometries*

`geom_icon` needs a new way to pack icons



# PGoG Grammar/*geometries*

`geom_icon` needs a new way to pack icons



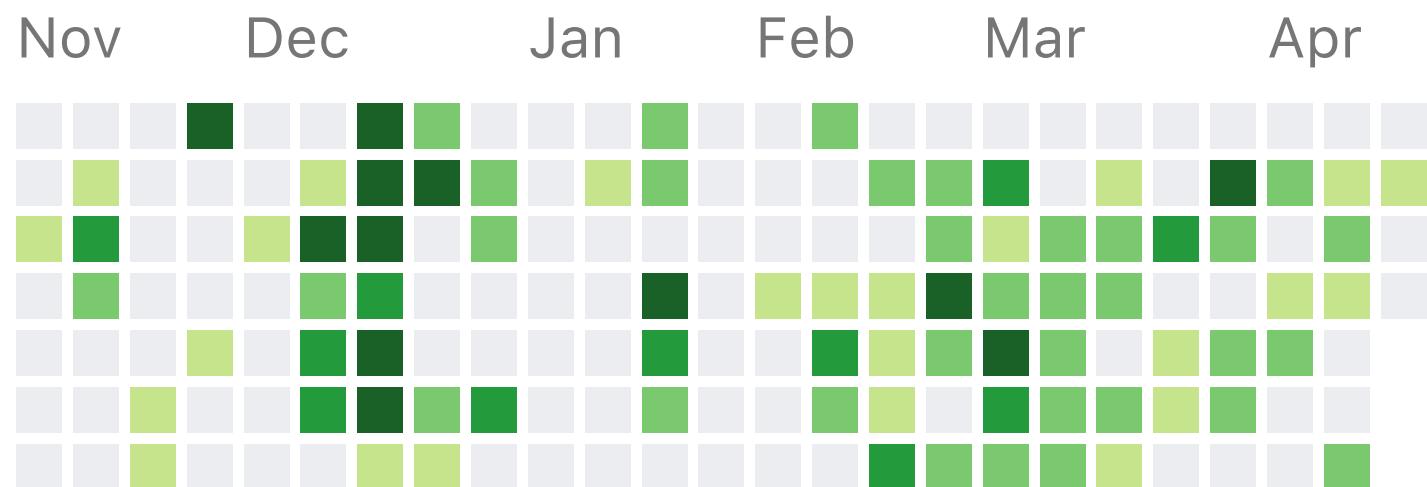
# Implementing the Grammar

## *Why in R*

- Grammar of graphics is implemented in ggplot2
- Metaprogramming features in R helps parsing
- PGoG grammar is transferrable

## *Current progress*

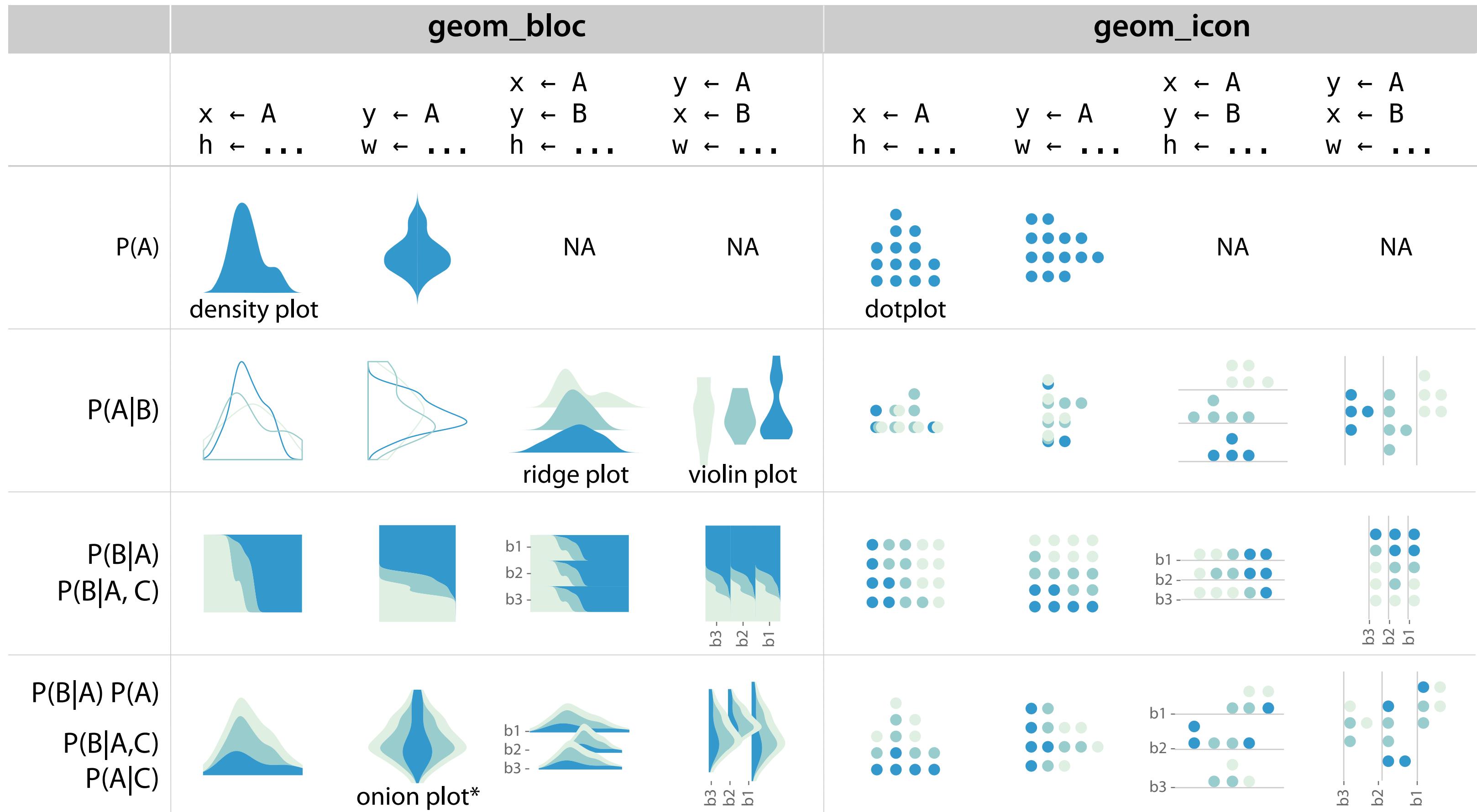
- geom\_bloc with discrete variables
- geom\_bloc with up to two continuous variables
- geom\_icon with up to two variables
- T0D0: aesthetics, partition bugs, parameters



# Evaluation of the Grammar

- Expressive?
- Generative?
- Cognitively ergonomic?

# Expressiveness of the grammar



# Generativeness from the combination of aesthetics



## Onion plot

`geom_bloc:`

`y ← mpg`

`width ← P(mpg) P(cyl|mpg)`

`direction ← both`

# Cognitive ergonomics

(Blackwell et al. 2001)

*Pro:*

Short edit distances

- *Low viscosity*
- *No premature commitment*

Close to probability expressions

*Con:*

Specifying probability expressions can be difficult

- *Hidden dependencies*
- *Error prone-ness*

# Cognitive ergonomics

*Pro:* Short edit distances, close to probability expressions

- *Low viscosity*
- *No premature commitment*

Existing ggplot2 packages

Changes

Syntax



---

```
geom_mosaic  
  x = cyl,  
      mpg*  
divider = hspine,  
      hspine
```

# Cognitive ergonomics

Pro: Short edit distances, close to probability expressions

- *Low viscosity*
- *No premature commitment*

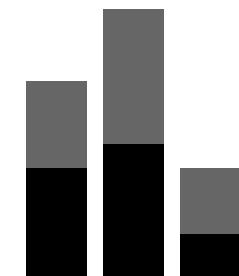
Existing ggplot2 packages

Changes

Syntax



geom:mosaic→bar  
+fill +y  
-divider



---

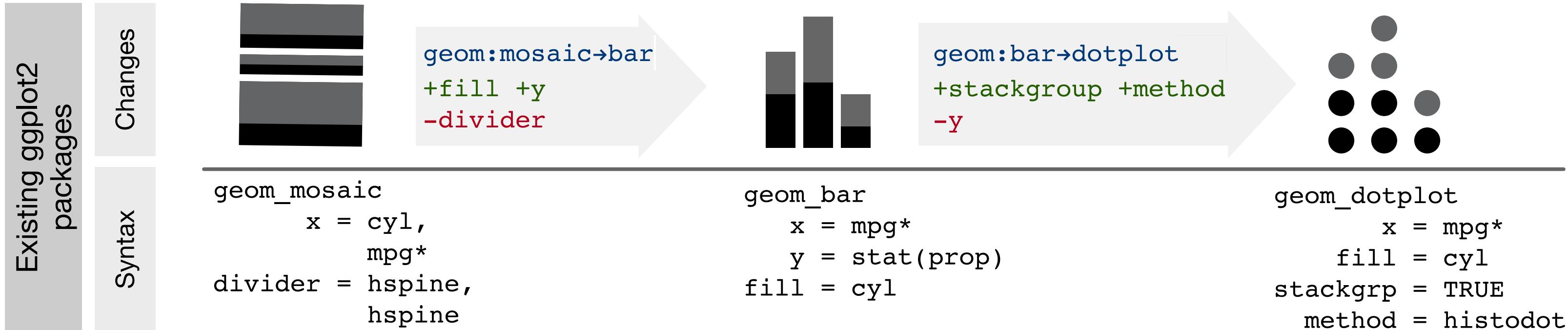
```
geom_mosaic
  x = cyl,
  mpg*
  divider = hspine,
  hspine
```

```
geom_bar
  x = mpg*
  y = stat(prop)
  fill = cyl
```

# Cognitive ergonomics

Pro: Short edit distances, close to probability expressions

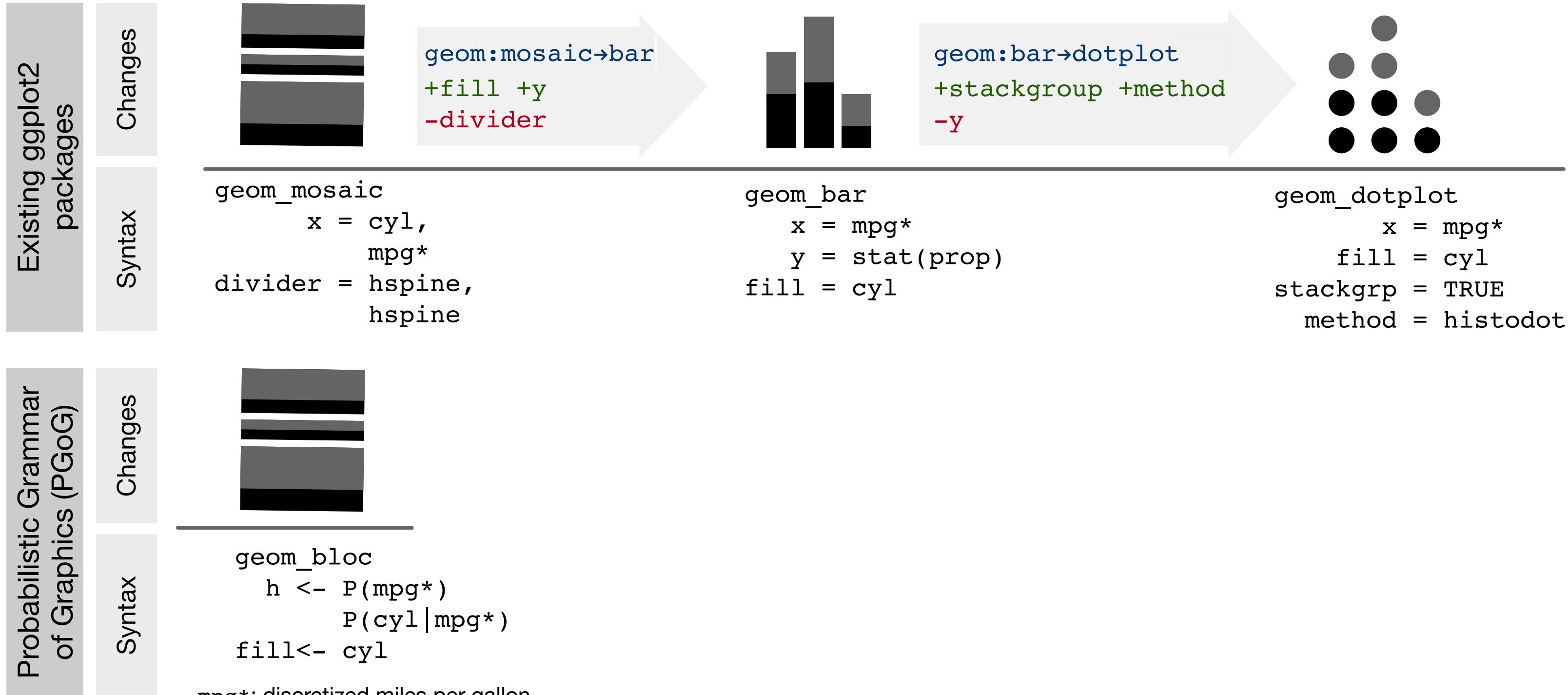
- *Low viscosity*
- *No premature commitment*



# Cognitive ergonomics

Pro: Short edit distances, close to probability expressions

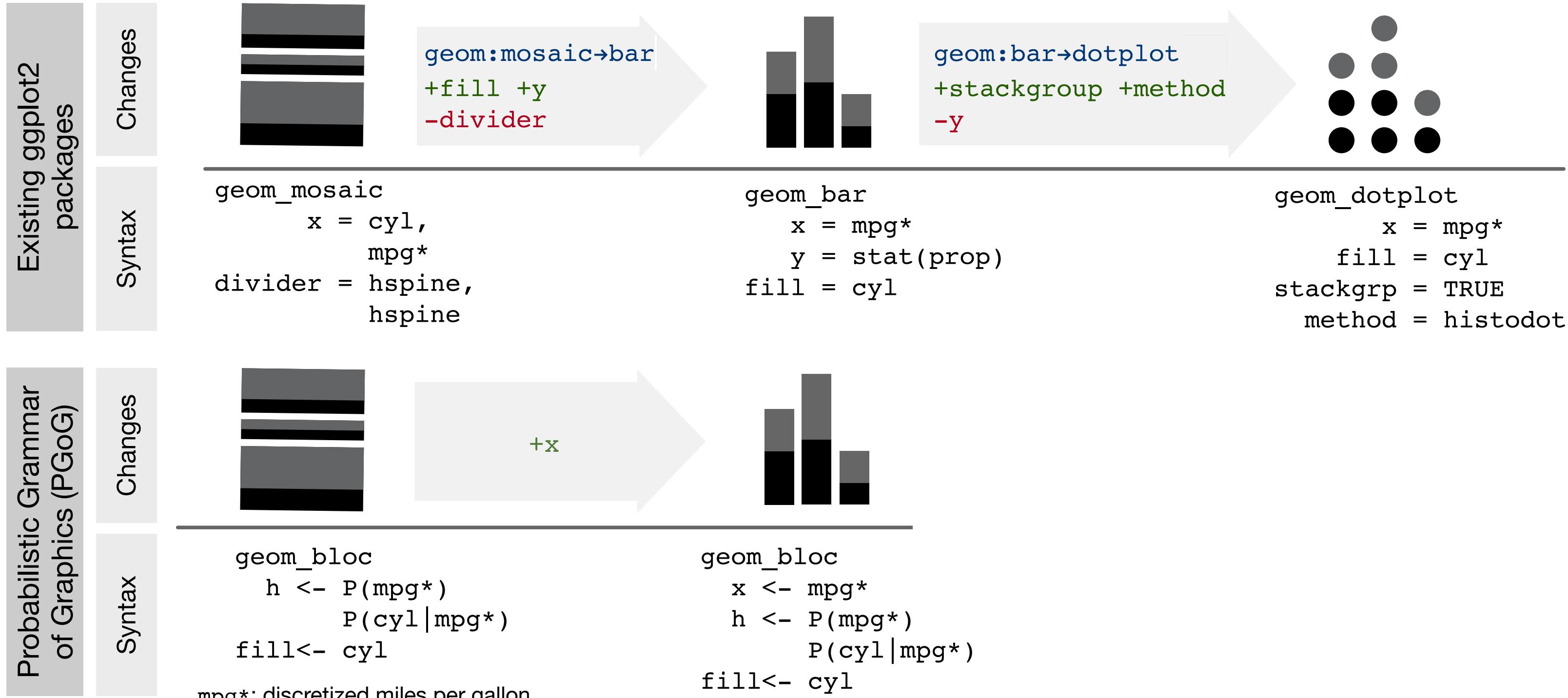
- *Low viscosity*
- *No premature commitment*



# Cognitive ergonomics

Pro: Short edit distances, close to probability expressions

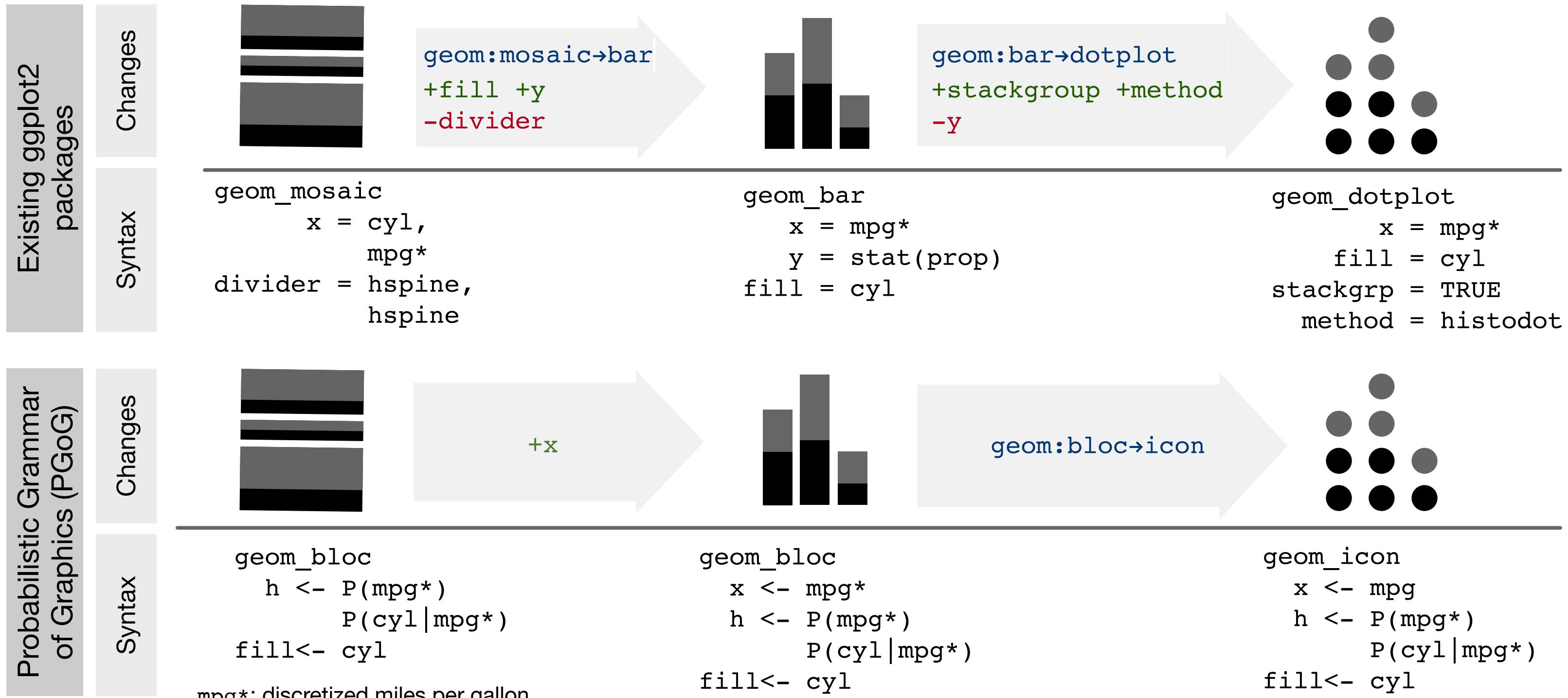
- *Low viscosity*
- *No premature commitment*



# Cognitive ergonomics

Pro: Short edit distances, close to probability expressions

- *Low viscosity*
- *No premature commitment*



# Cognitive ergonomics

*Con:* specifying probability expressions can be difficult

- *Hidden dependencies*
- *Error prone-ness*

Math

$$P(\text{mpg} | \text{cyl}) \quad P(\text{mpg}) ?$$



$$\mathbf{P}(\text{mpg} | \text{cyl}) \quad \mathbf{P}(\text{cyl}) = \mathbf{P}(\text{mpg}, \text{cyl})$$

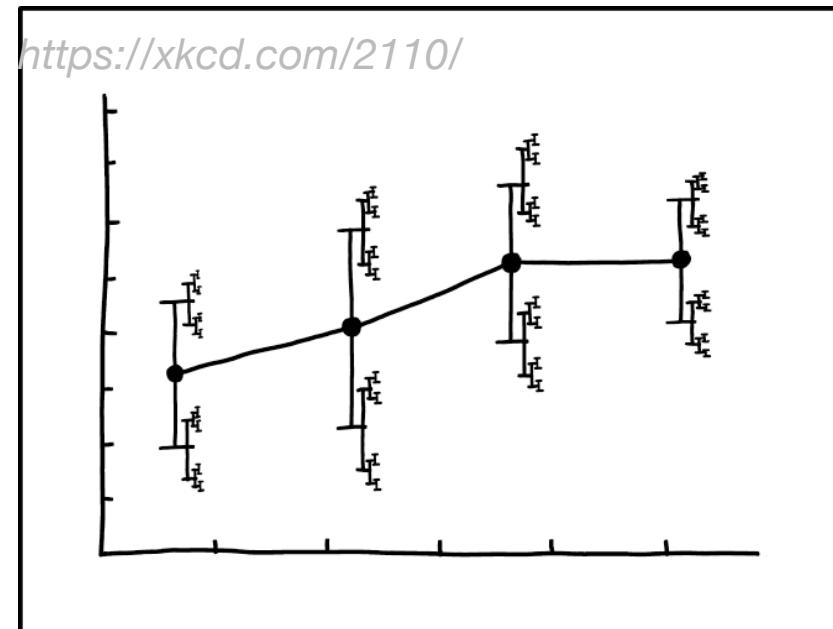
# Cognitive ergonomics

Con: specifying probability expressions can be difficult

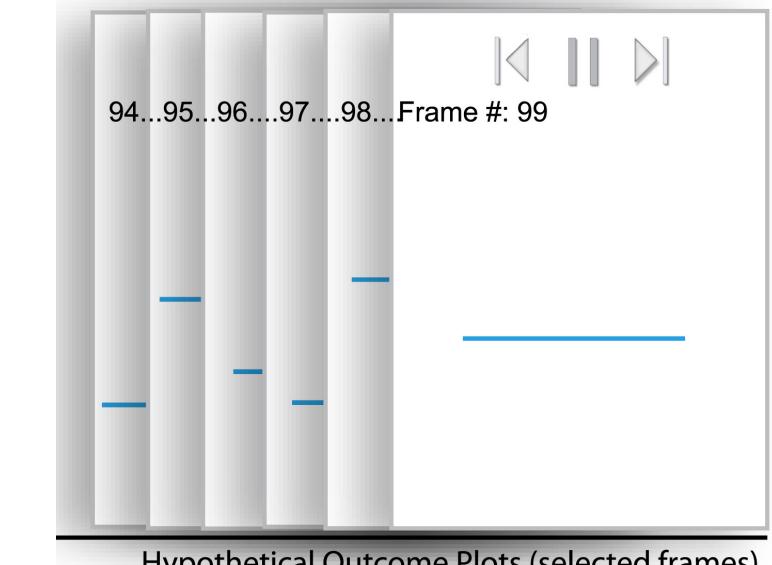
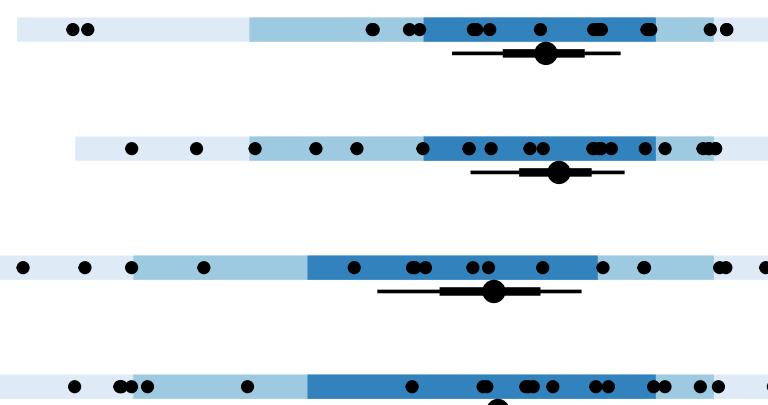
- *Hidden dependencies*
- *Error prone-ness*

|            |                                                                          |
|------------|--------------------------------------------------------------------------|
|            | $P(\text{mpg}   \text{cyl}) \ P(\text{mpg}) ?$                           |
| Math       | $P(\text{mpg}   \text{cyl}) \ P(\text{cyl}) = P(\text{mpg}, \text{cyl})$ |
| Coord aes  | $x \leftarrow \text{mpg}$                                                |
| Prob aes   | $\text{height} \leftarrow P(\text{cyl}) \ P(\text{mpg}   \text{cyl})$    |
| Visual aes | $\text{fill} \leftarrow \text{cyl}$                                      |

# Future work: more uncertainty vizes & systemization

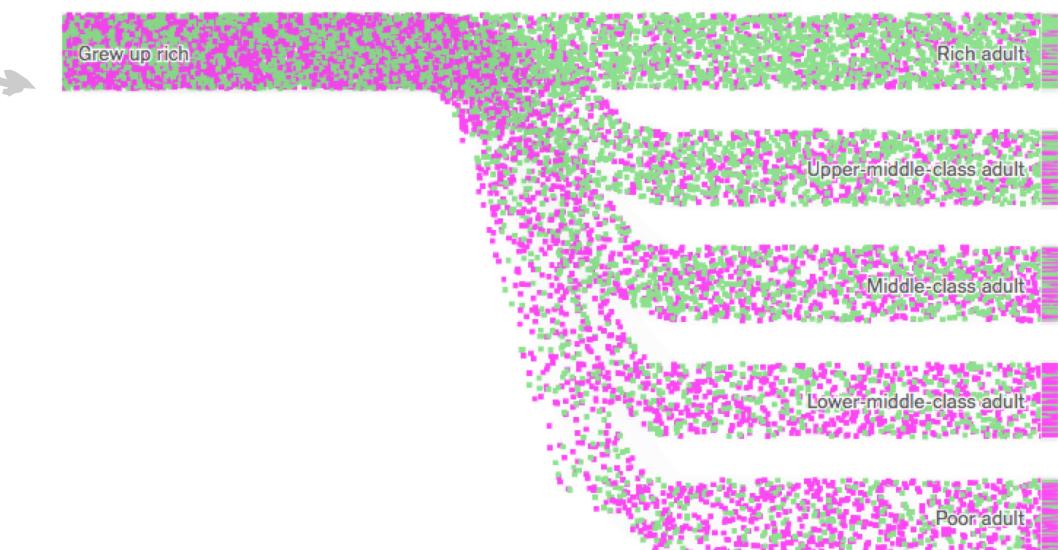


- Uncertainty sources: aleatory or epistemic
- Data structure: hierarchical, sequential, etc.
- Summary statistics, confidence intervals, etc.
- Visualization techniques such as linking



(Hullman, Resnick, and Adar 2015)

[https://www.nytimes.com/interactive/2018/03/27/up-  
shot/make-your-own-mobility-animation.html](https://www.nytimes.com/interactive/2018/03/27/up-shot/make-your-own-mobility-animation.html)



# Conclusions

PGoG is a visualization grammar for **probabilistic visualizations**, treating probability distributions as first-class citizens. It **shifts our thinking** about specifications for probabilistic visualizations and could facilitate **uncertainty** communication in the future.