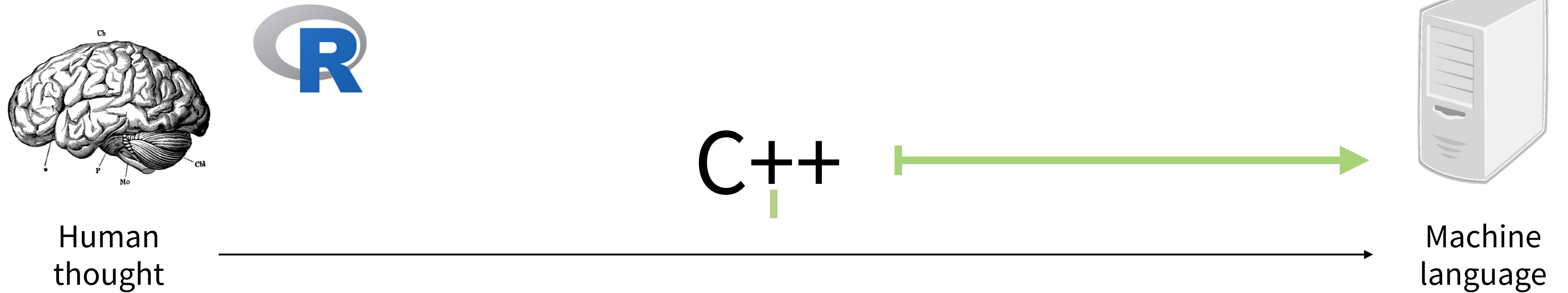


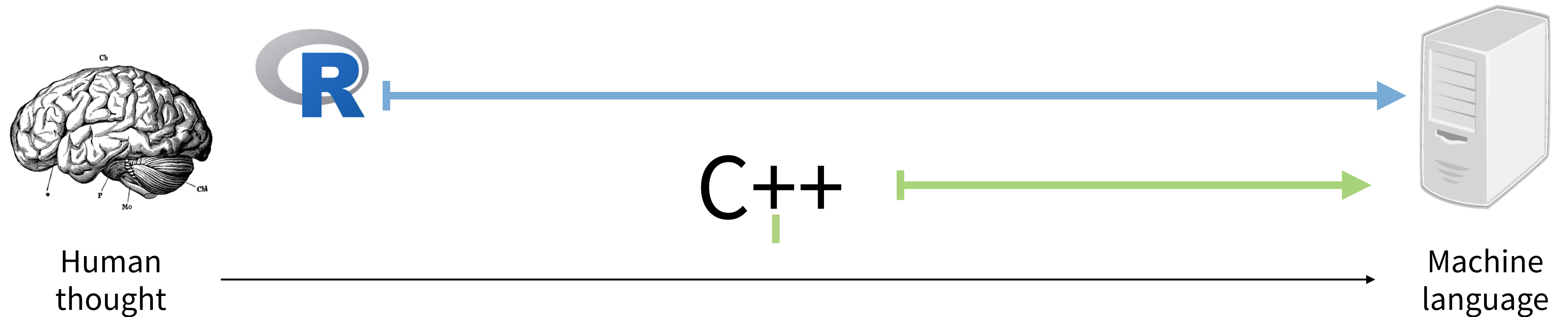
ggplot2

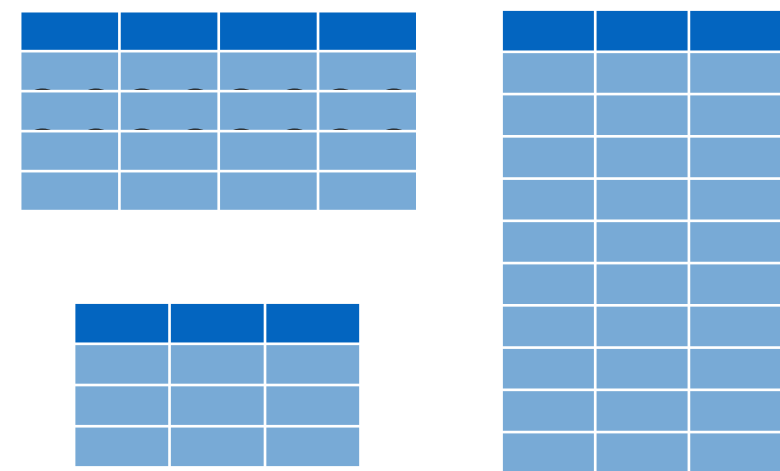


R - A computer language for scientists

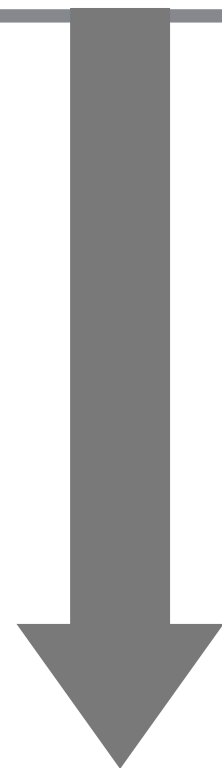
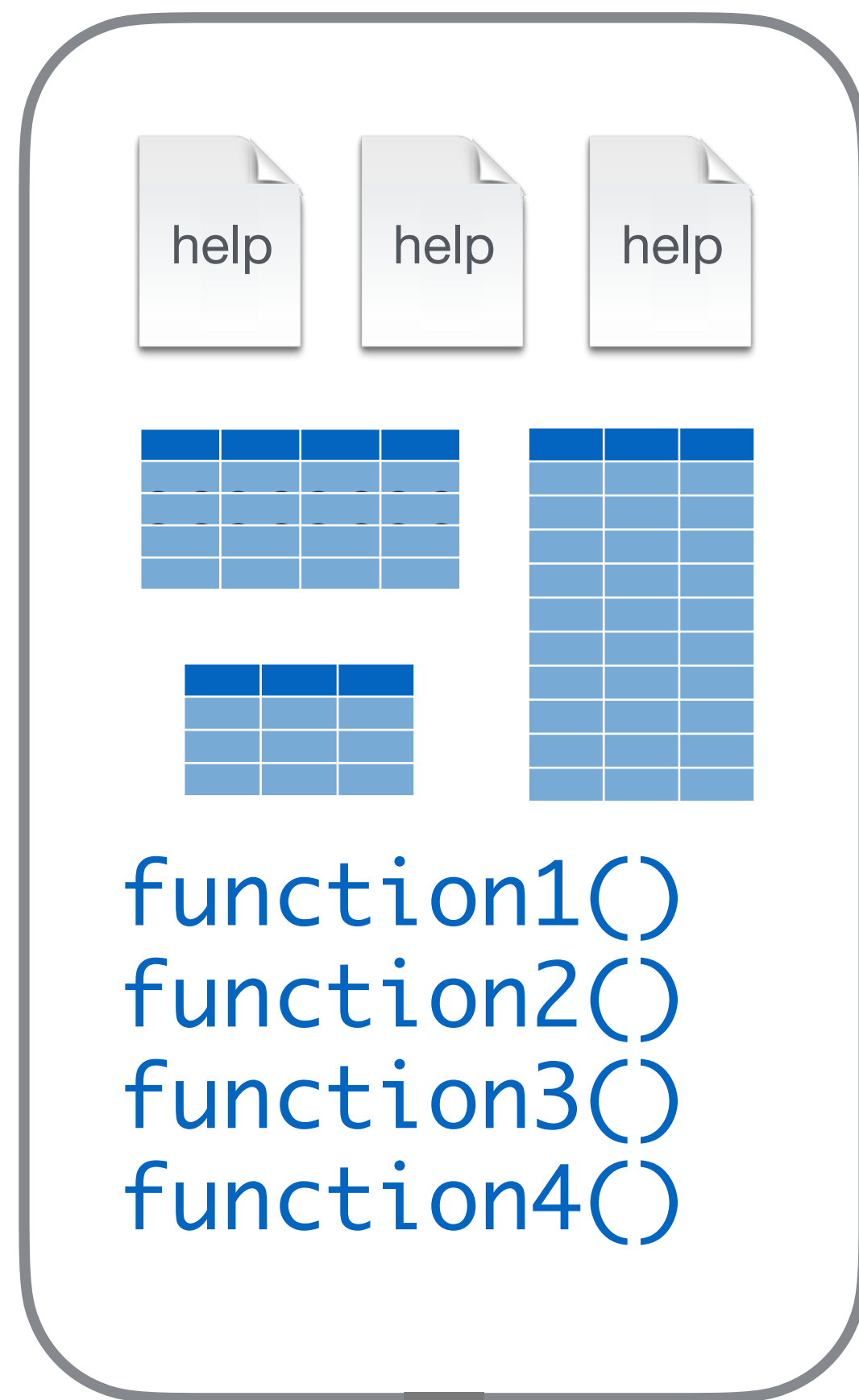


R - A computer language for scientists

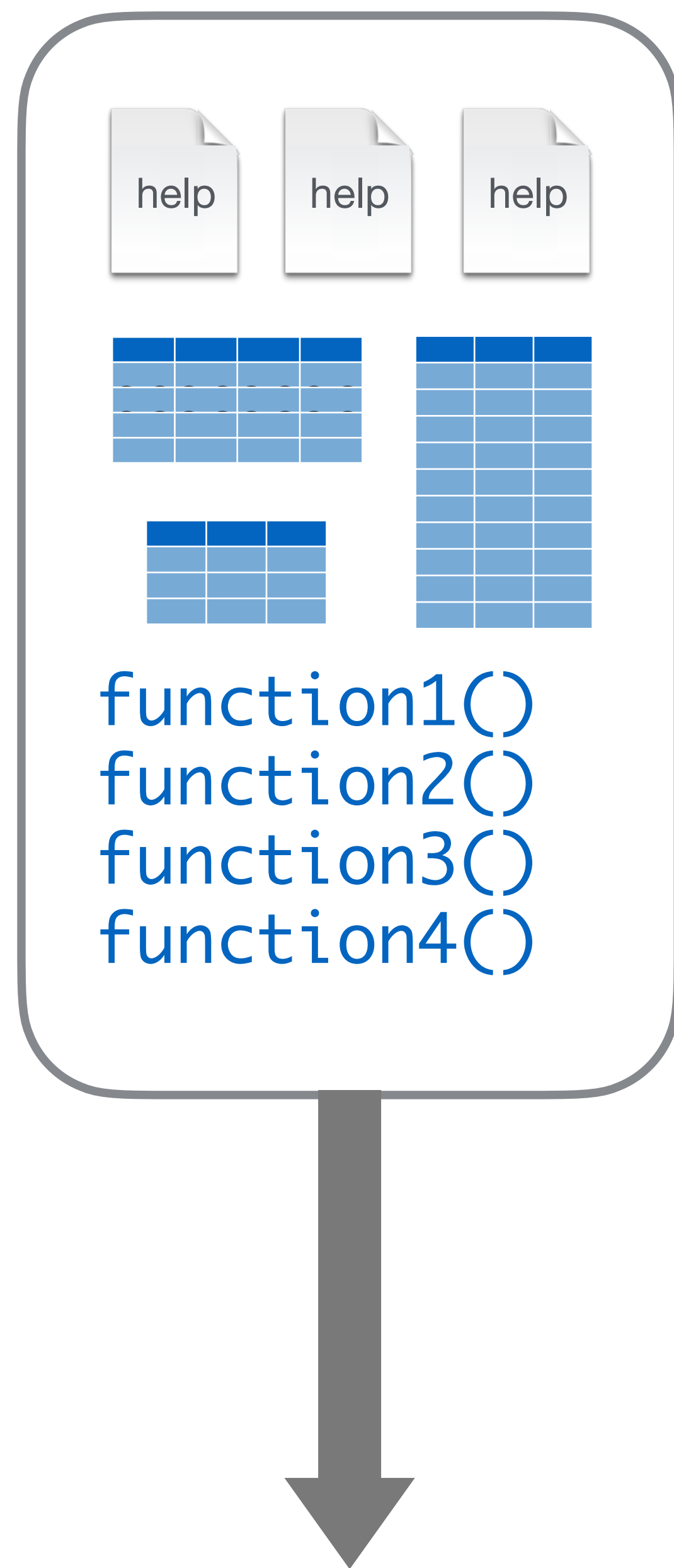




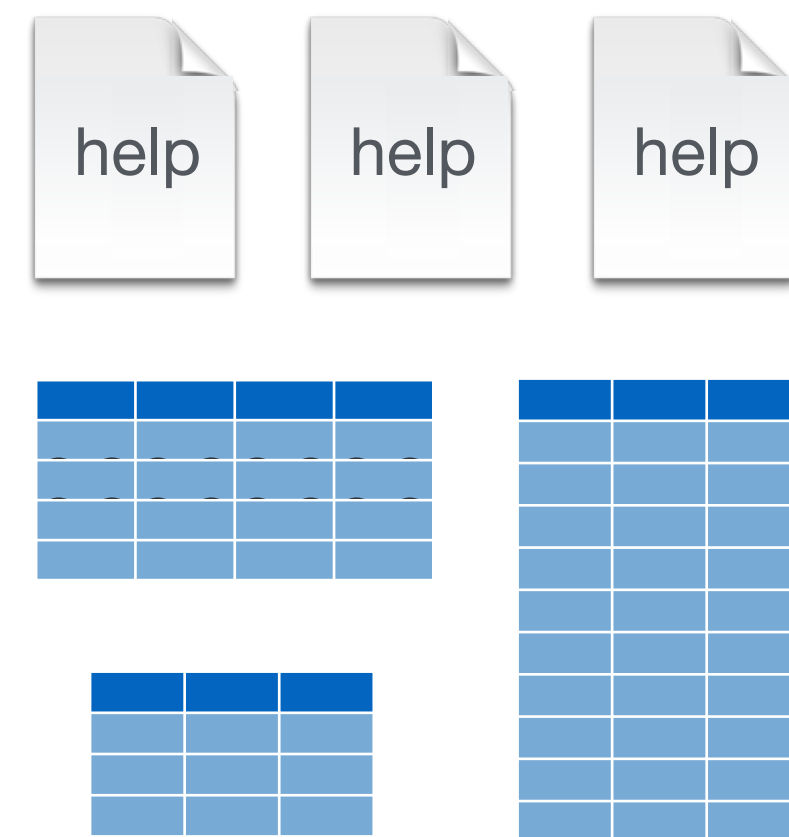
function1()
function2()
function3()
function4()



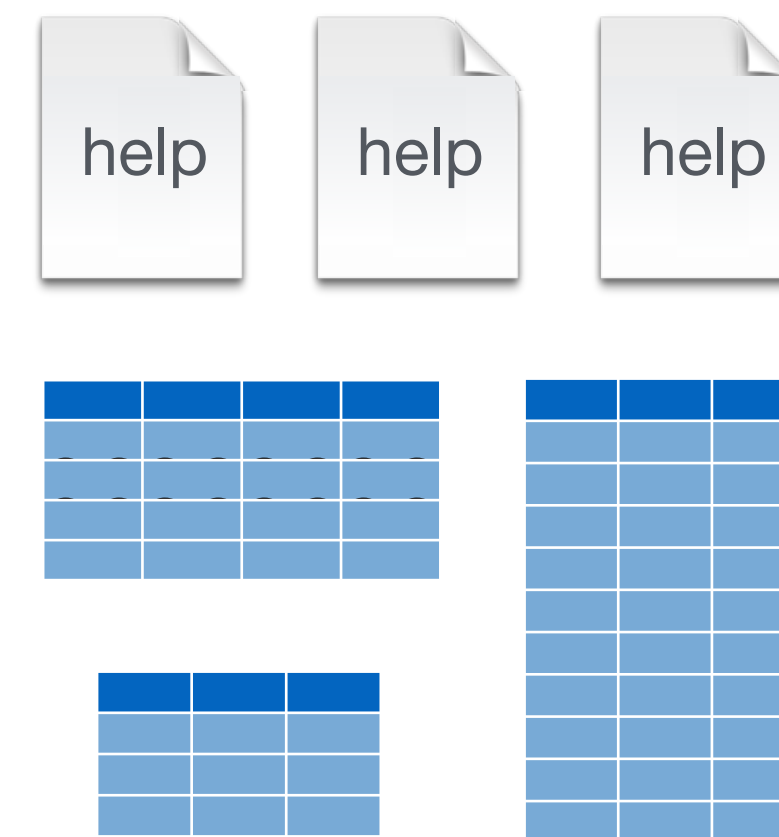
Base R



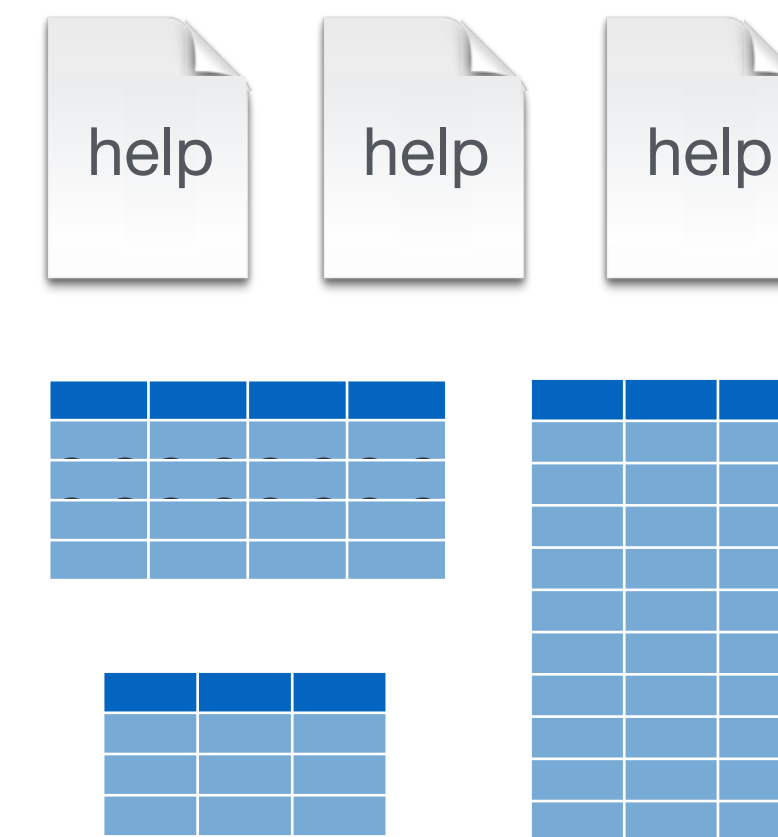
Base R



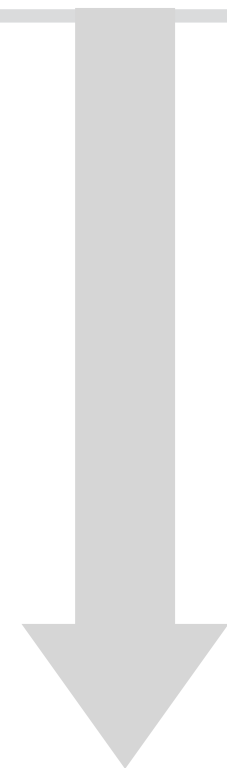
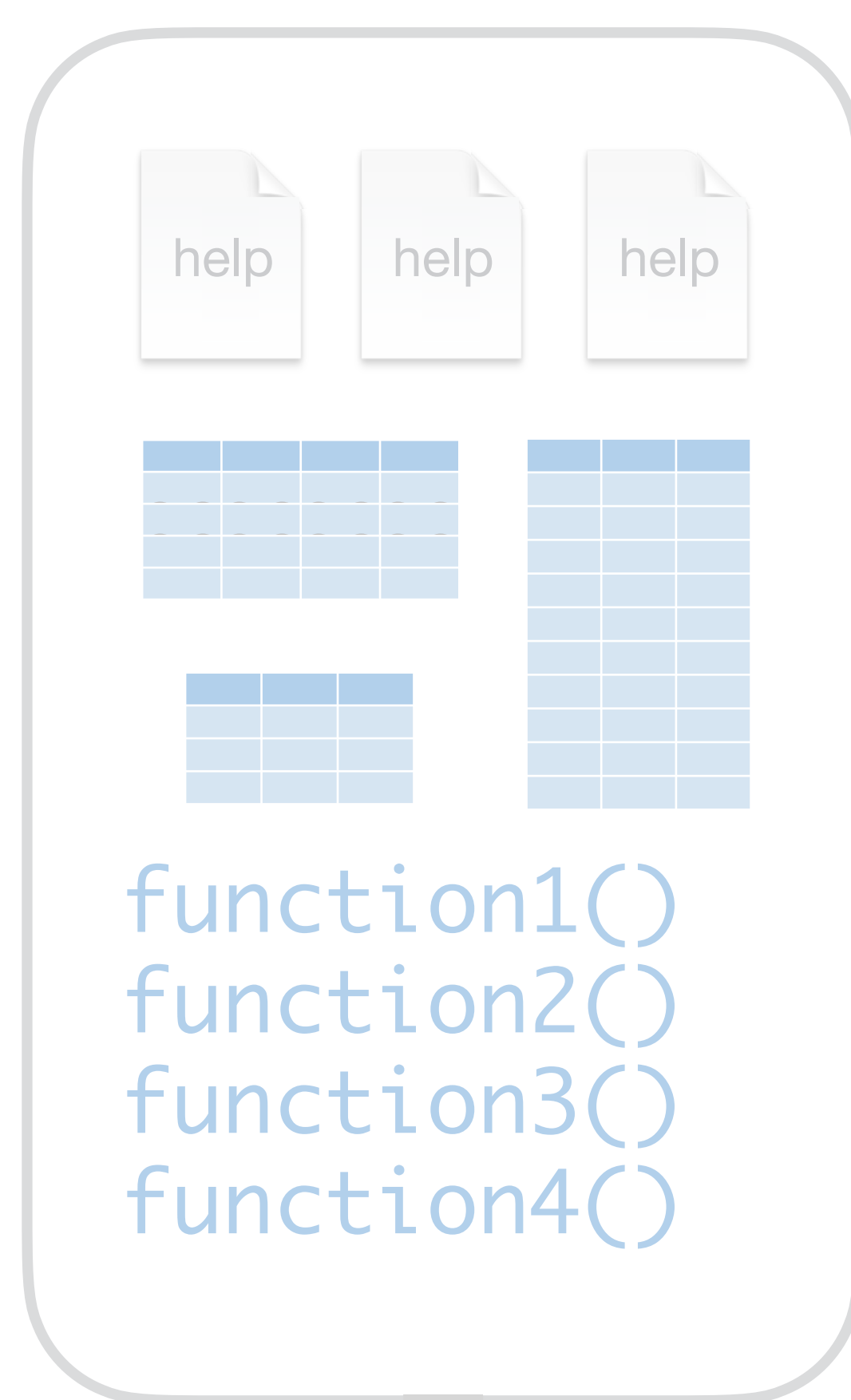
`function5()`
`function6()`
`function7()`
`function8()`



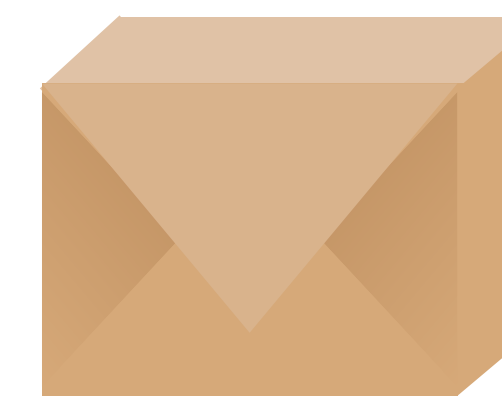
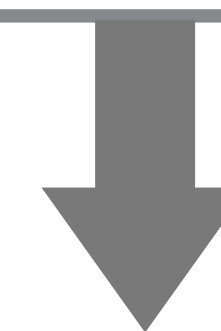
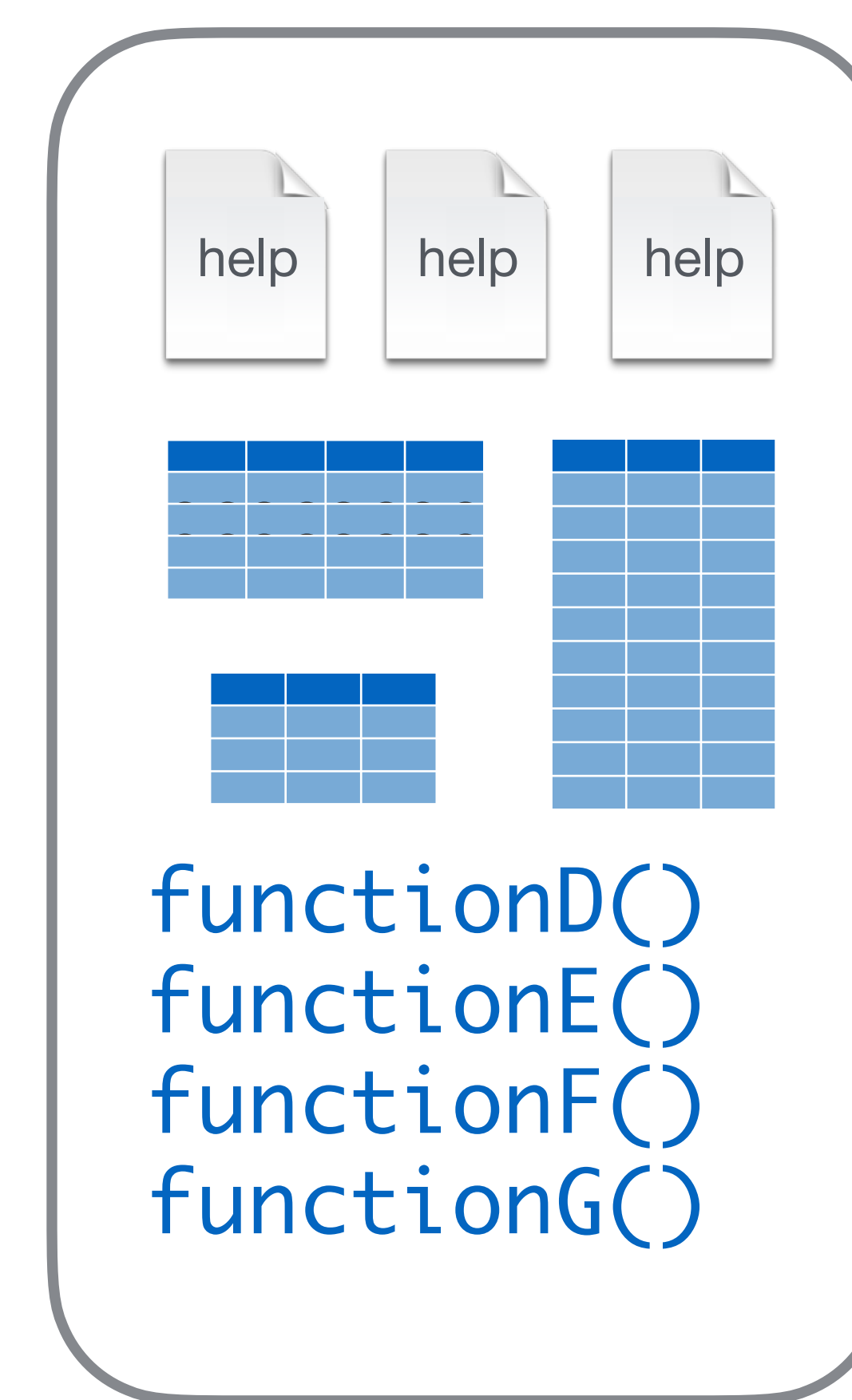
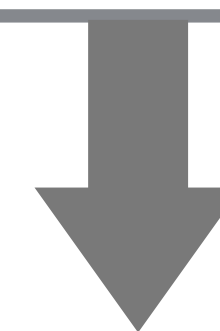
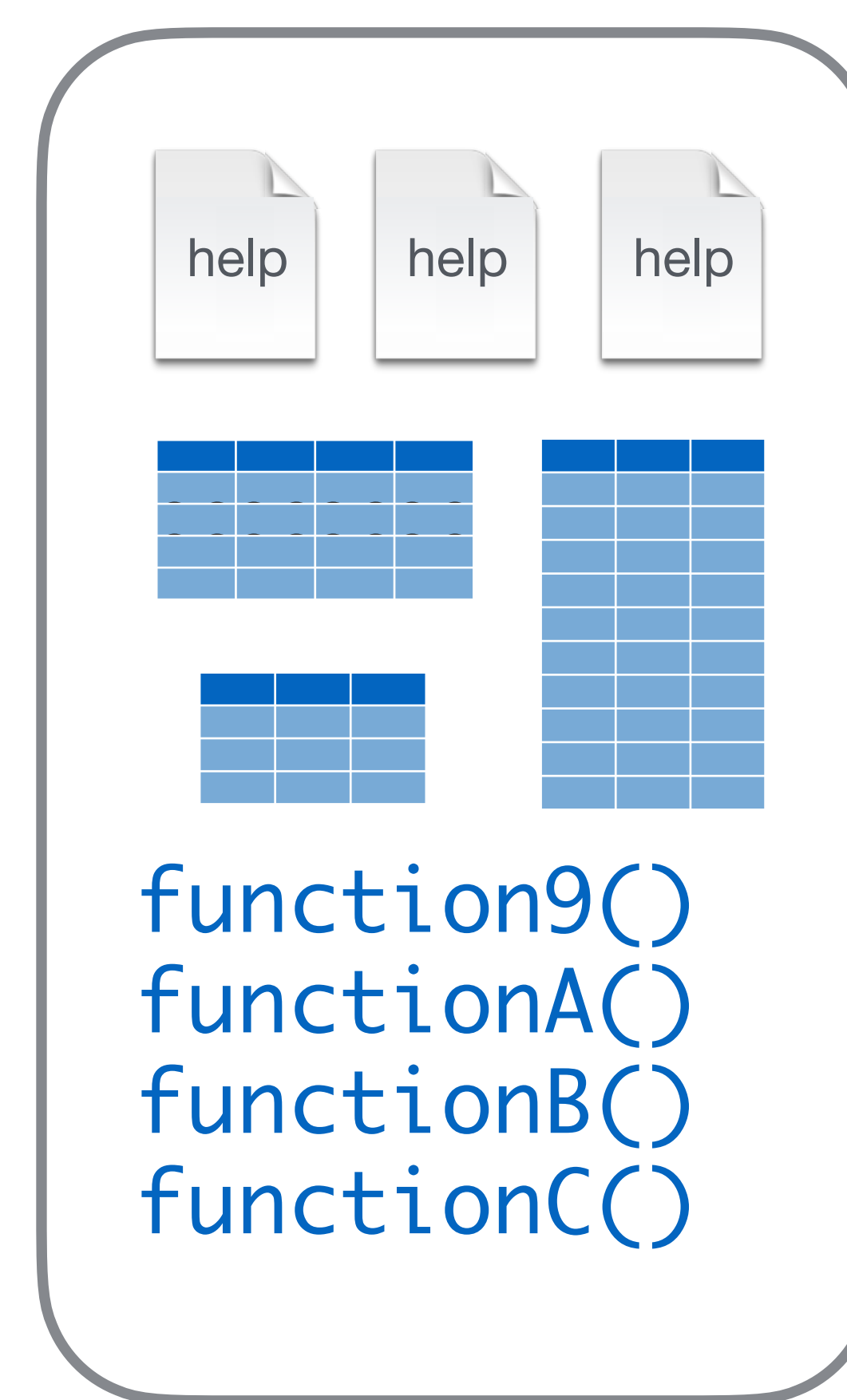
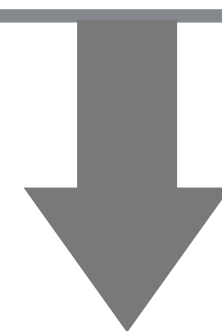
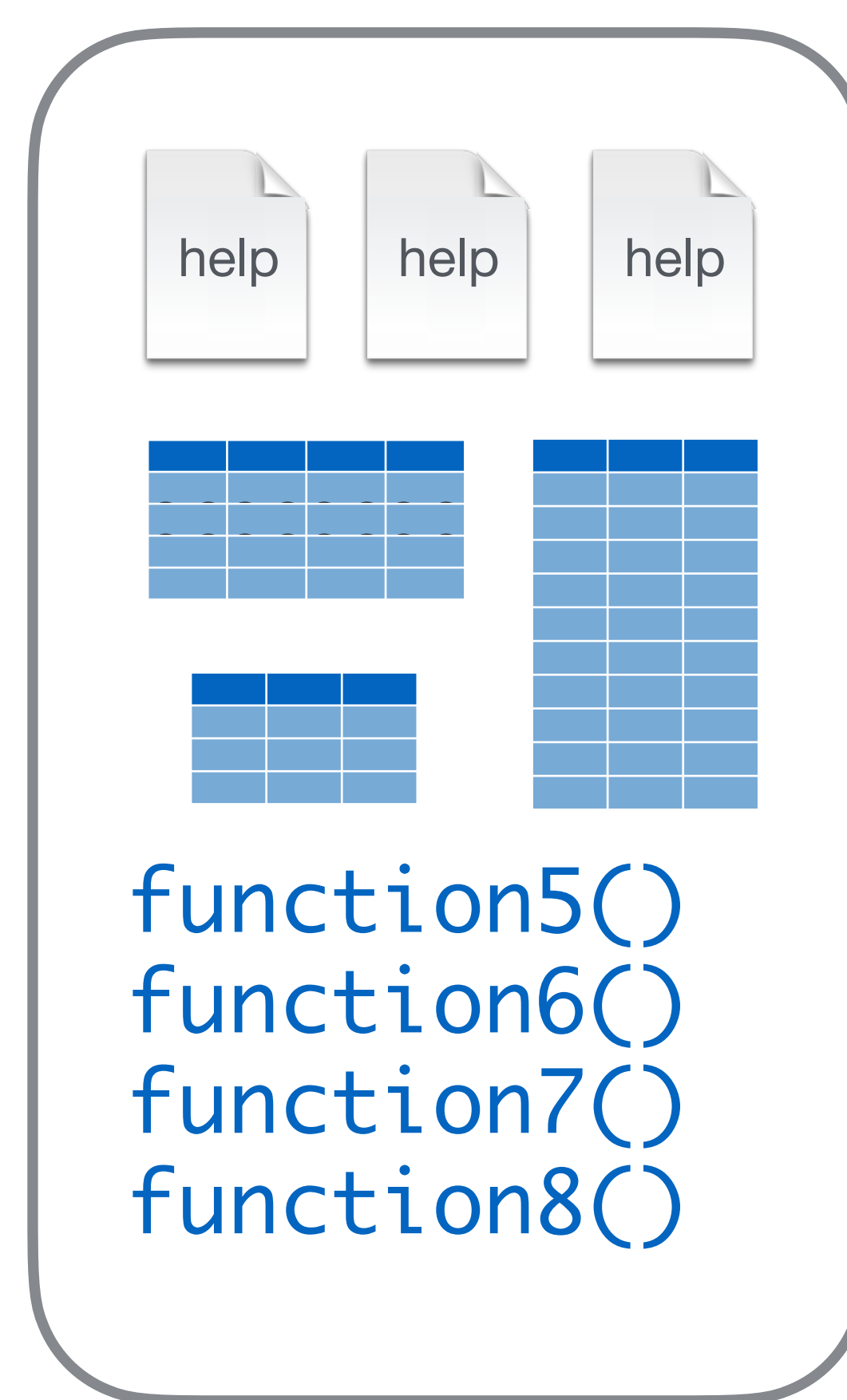
`function9()`
`functionA()`
`functionB()`
`functionC()`



`functionD()`
`functionE()`
`functionF()`
`functionG()`



Base R



R Packages

Available CRAN Packages By Name

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

A3	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
AalenJohansen	Conditional Aalen-Johansen Estimation
AATtools	Reliability and Scoring Routines for the Approach-Avoidance Task
ABACUS	Apps Based Activities for Communicating and Understanding Statistics
abasequence	Coding 'ABA' Patterns for Sequence Data
abbreviate	Readable String Abbreviation
abbyyR	Access to Abbyy Optical Character Recognition (OCR) API
abc	Tools for Approximate Bayesian Computation (ABC)
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)
ABC.RAP	Array Based CpG Region Analysis Pipeline
ABCanalysis	Computed ABC Analysis
abclass	Angle-Based Large-Margin Classifiers
ABCoptim	Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2	Approximate Bayesian Computational Model for Estimating P2
abcrf	Approximate Bayesian Computation via Random Forests
abcrlda	Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis
abctools	Tools for ABC Analyses
abd	The Analysis of Biological Data
abdiv	Alpha and Beta Diversity Measures
abe	Augmented Backward Elimination
abess	Fast Best Subset Selection
abglasso	Adaptive Bayesian Graphical Lasso

Using packages

Using packages

1

```
install.packages("foo")
```

Downloads files to "computer"

Using packages

1

```
install.packages("foo")
```

Downloads files to "computer"

2

```
library(foo)
```

Loads package

Using packages

1

```
install.packages("foo")
```

Downloads files to "computer"

1 x per "computer"

2

```
library(foo)
```

Loads package

Using packages

1

```
install.packages("foo")
```

Downloads files to "computer"

1 x per "computer"

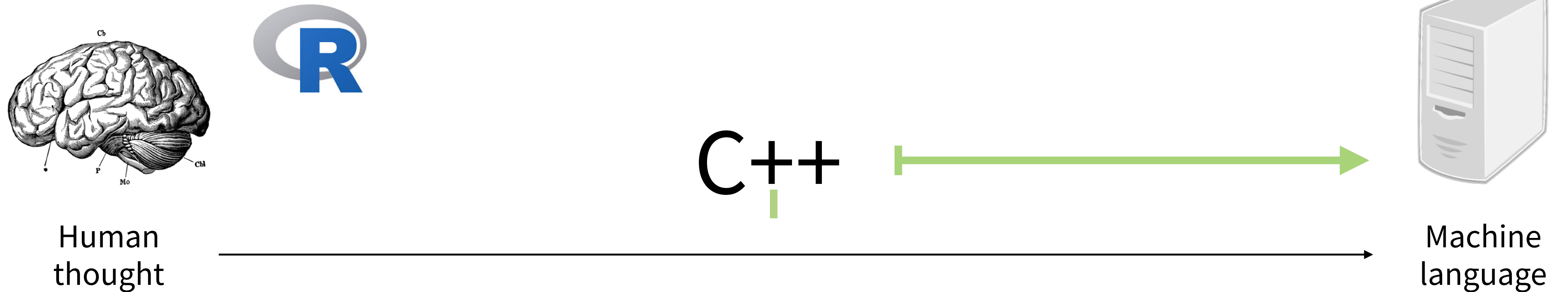
2

```
library(foo)
```

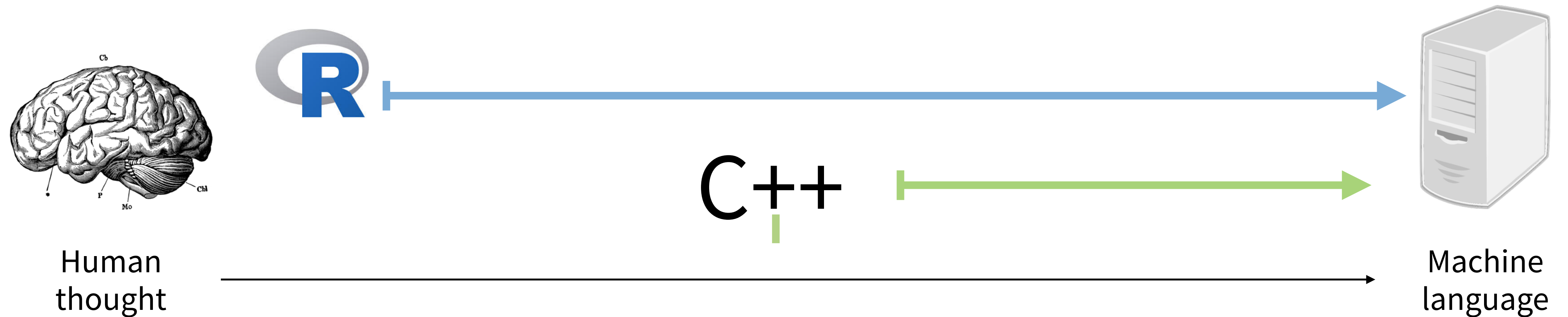
Loads package

1 x per R Session

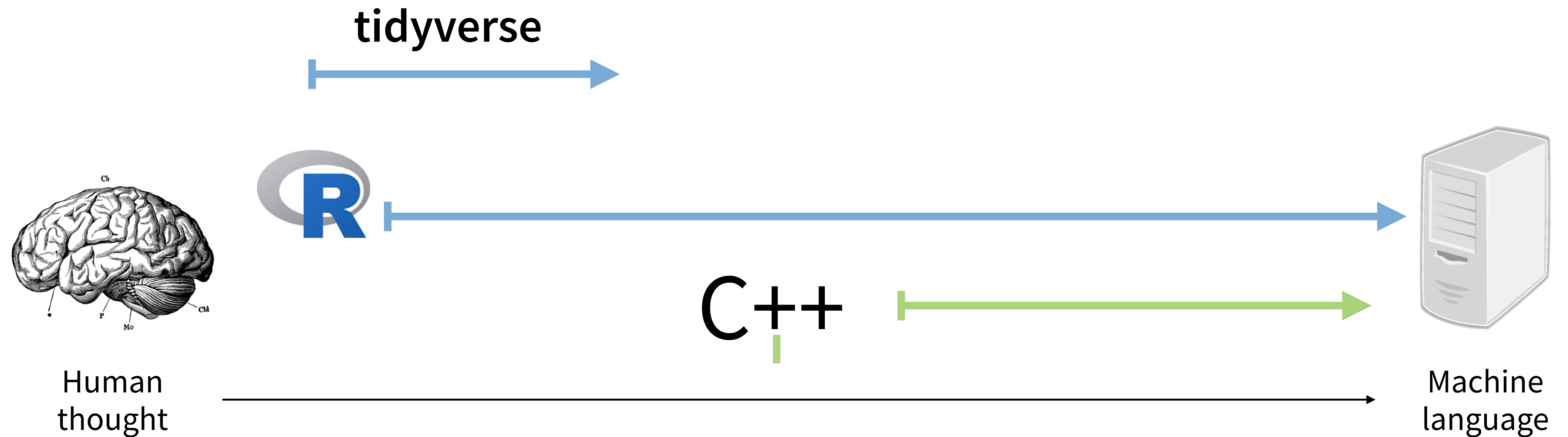
The tidyverse - A set of R packages to unify some data science tasks



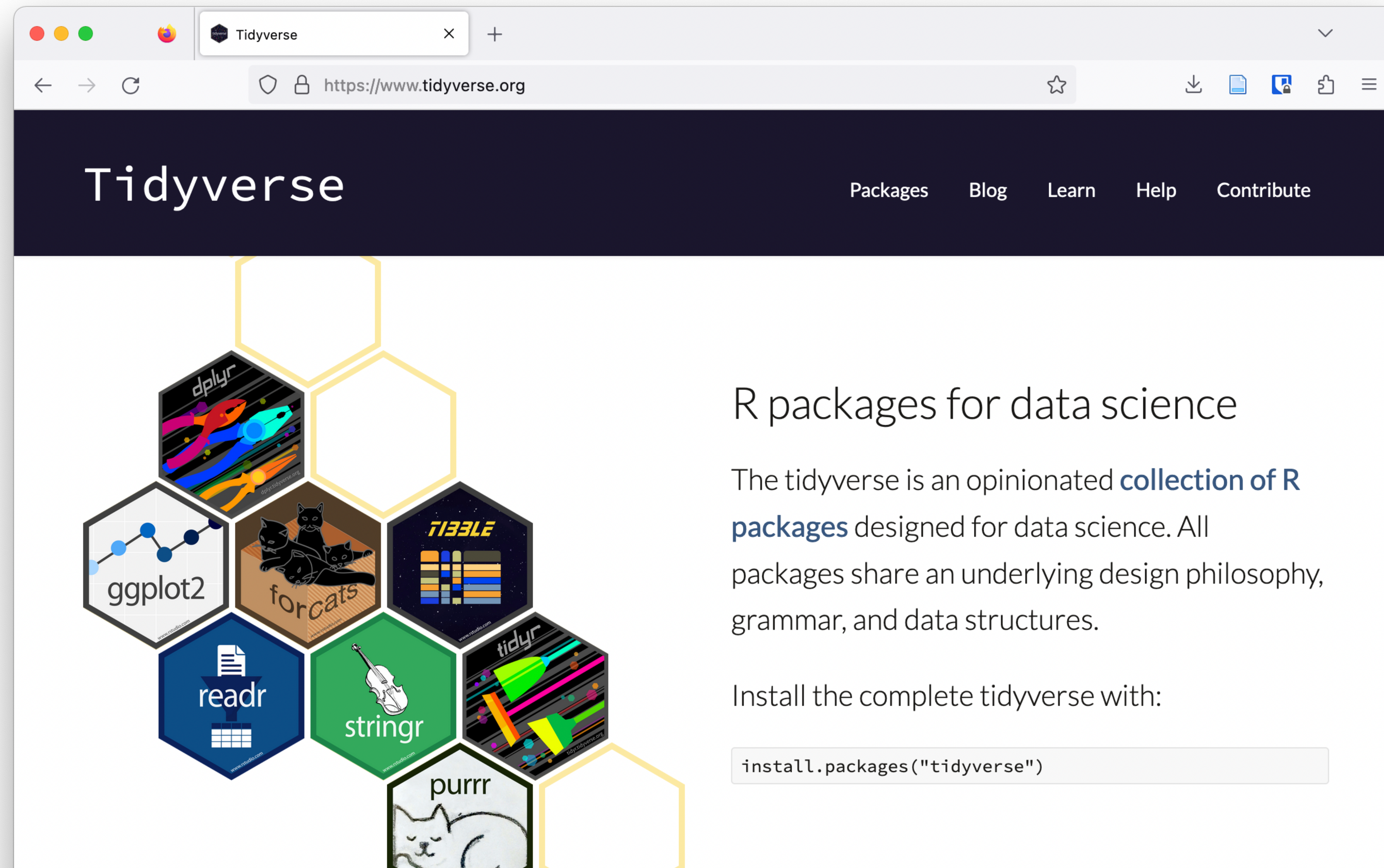
The tidyverse - A set of R packages to unify some data science tasks



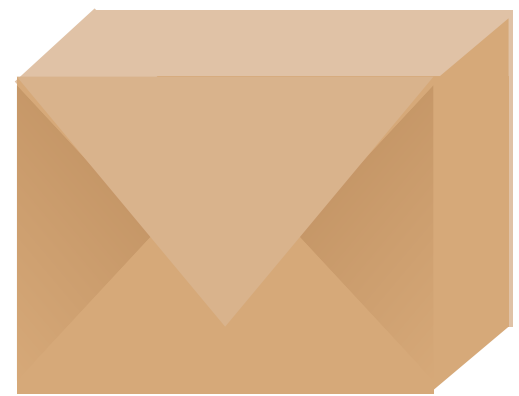
The tidyverse - A set of R packages to unify some data science tasks



tidyverse.org



tidyverse



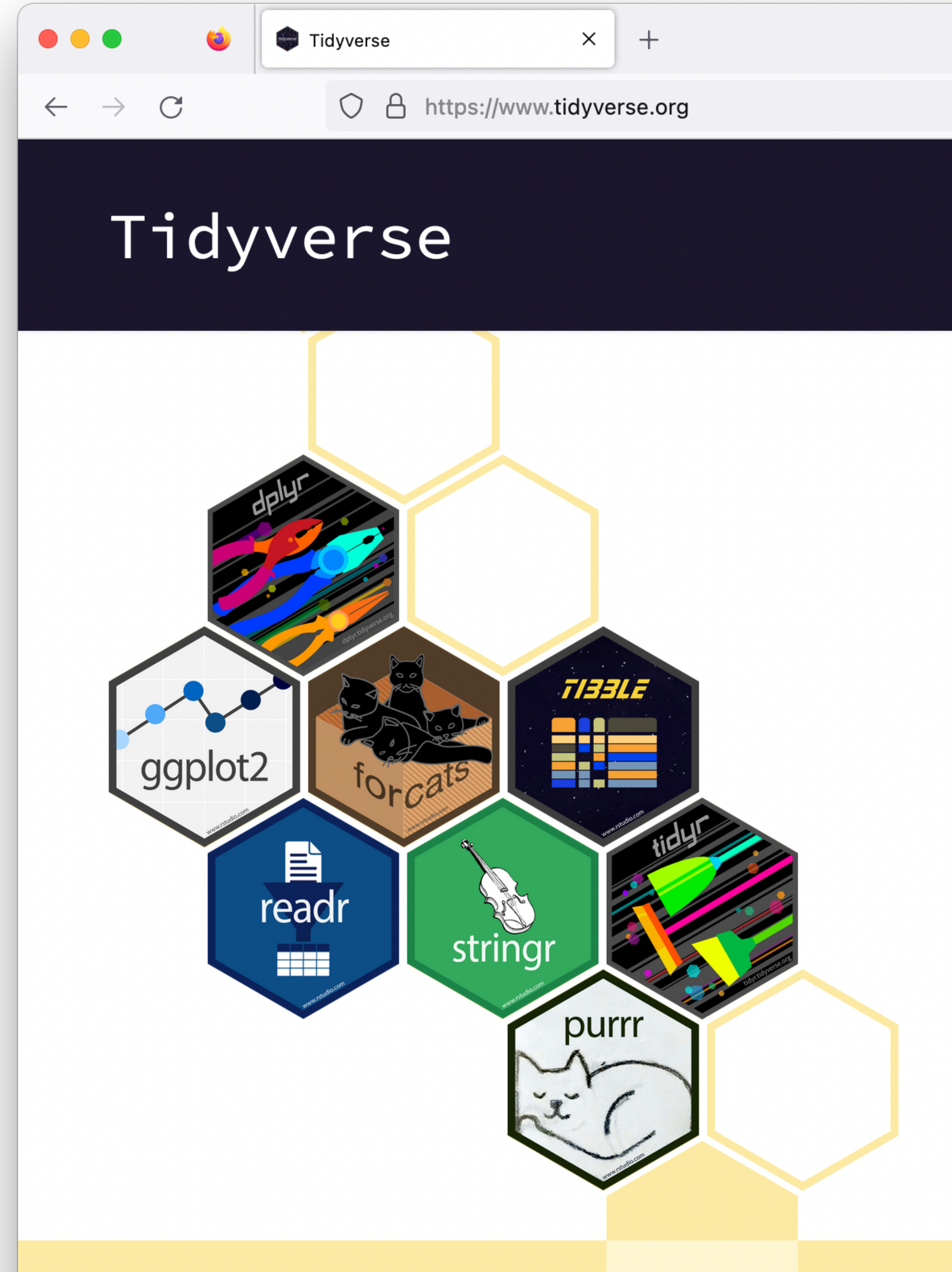
An R package that serves as a short cut for installing and loading the components of the tidyverse.

```
library(tidyverse)
```

```
install.packages("tidyverse")
```

does the equivalent of

```
install.packages("ggplot2")  
install.packages("dplyr")  
install.packages("tidyr")  
install.packages("readr")  
install.packages("purrr")  
install.packages("tibble")  
install.packages("hms")  
install.packages("stringr")  
install.packages("lubridate")  
install.packages("forcats")  
install.packages("DBI")  
install.packages("haven")  
install.packages("httr")  
install.packages("jsonlite")  
install.packages("readxl")  
install.packages("rvest")  
install.packages("xml2")  
install.packages("modelr")  
install.packages("broom")
```



```
install.packages("tidyverse")
```

does the equivalent of

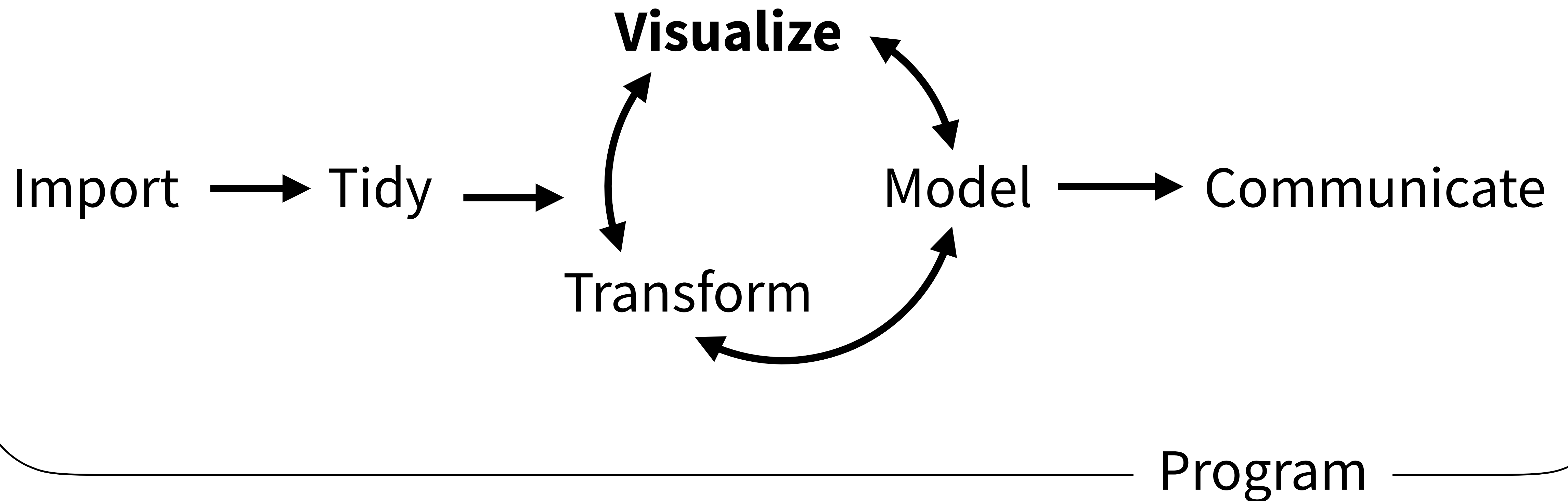
```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
install.packages("lubridate")
install.packages("forcats")
install.packages("hms")
install.packages("DBI")
install.packages("haven")
install.packages("httr")
install.packages("jsonlite")
install.packages("readxl")
install.packages("rvest")
install.packages("xml2")
install.packages("modelr")
install.packages("broom")
```

```
library(tidyverse)
```

does the equivalent of

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(readr)
library(purrr)
library(tibble)
library(stringr)
library(lubridate)
library(forcats)
```

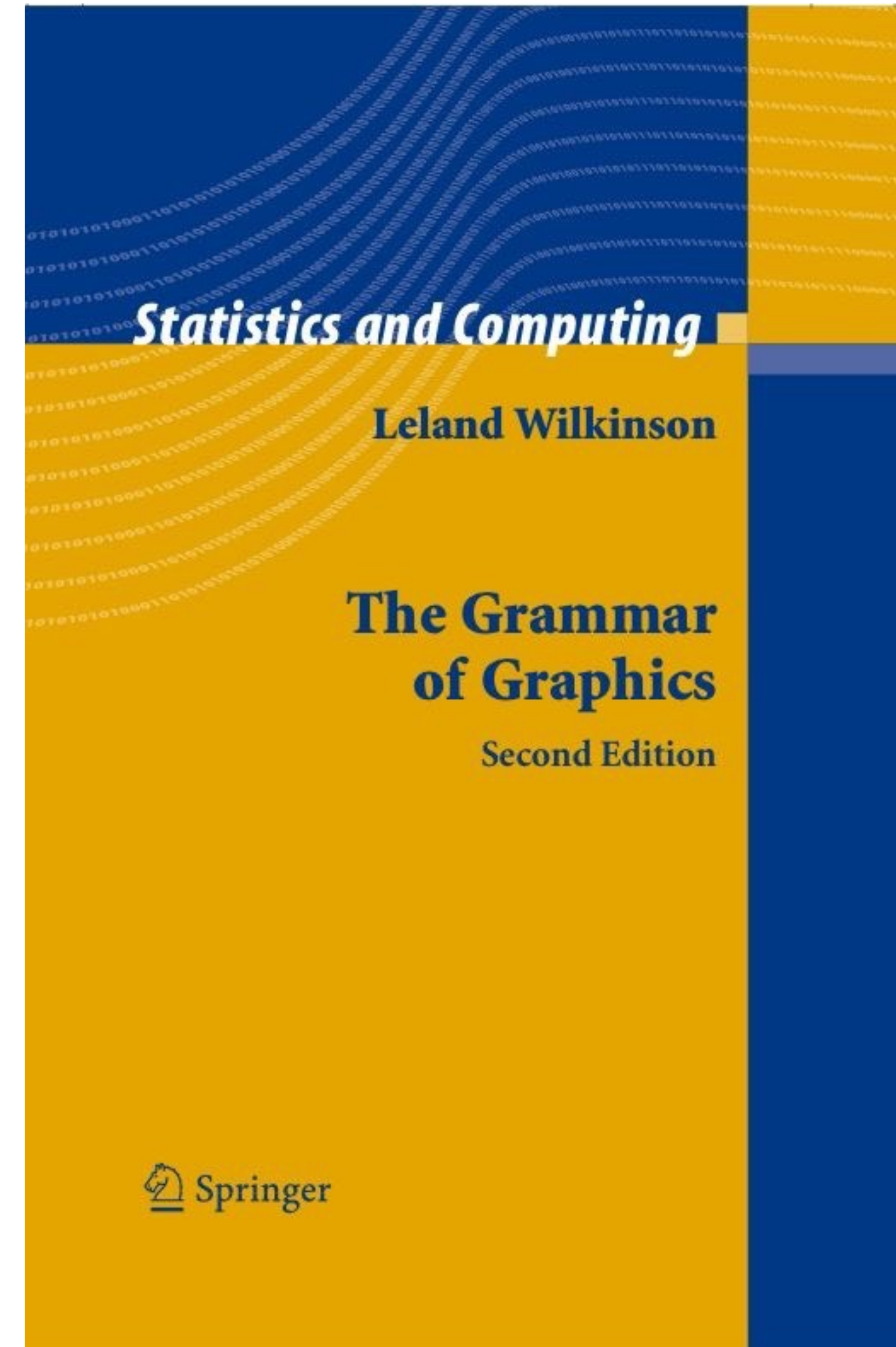
(Applied) Data Science



Grammar of Graphics

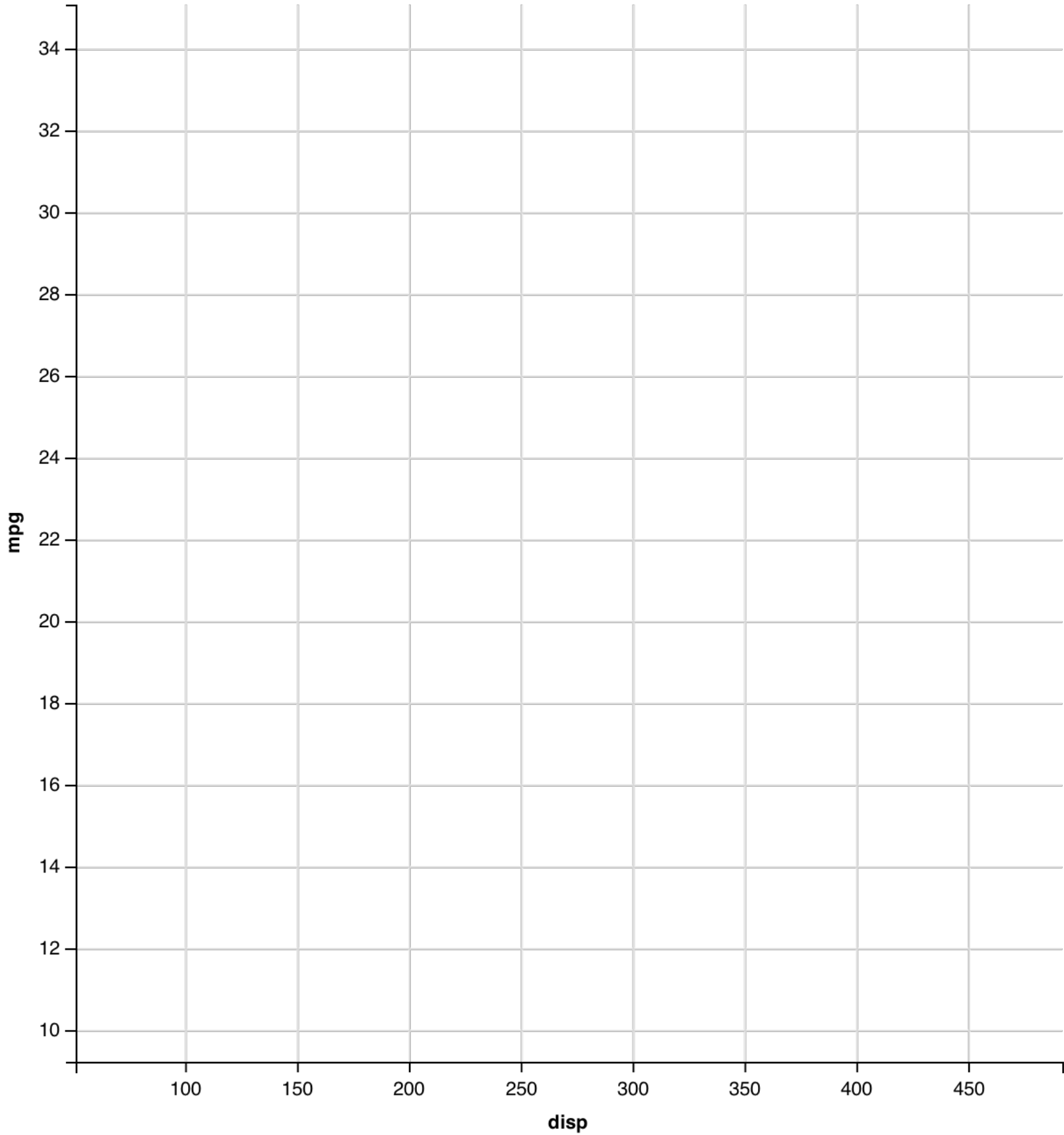


The Grammar of Graphics



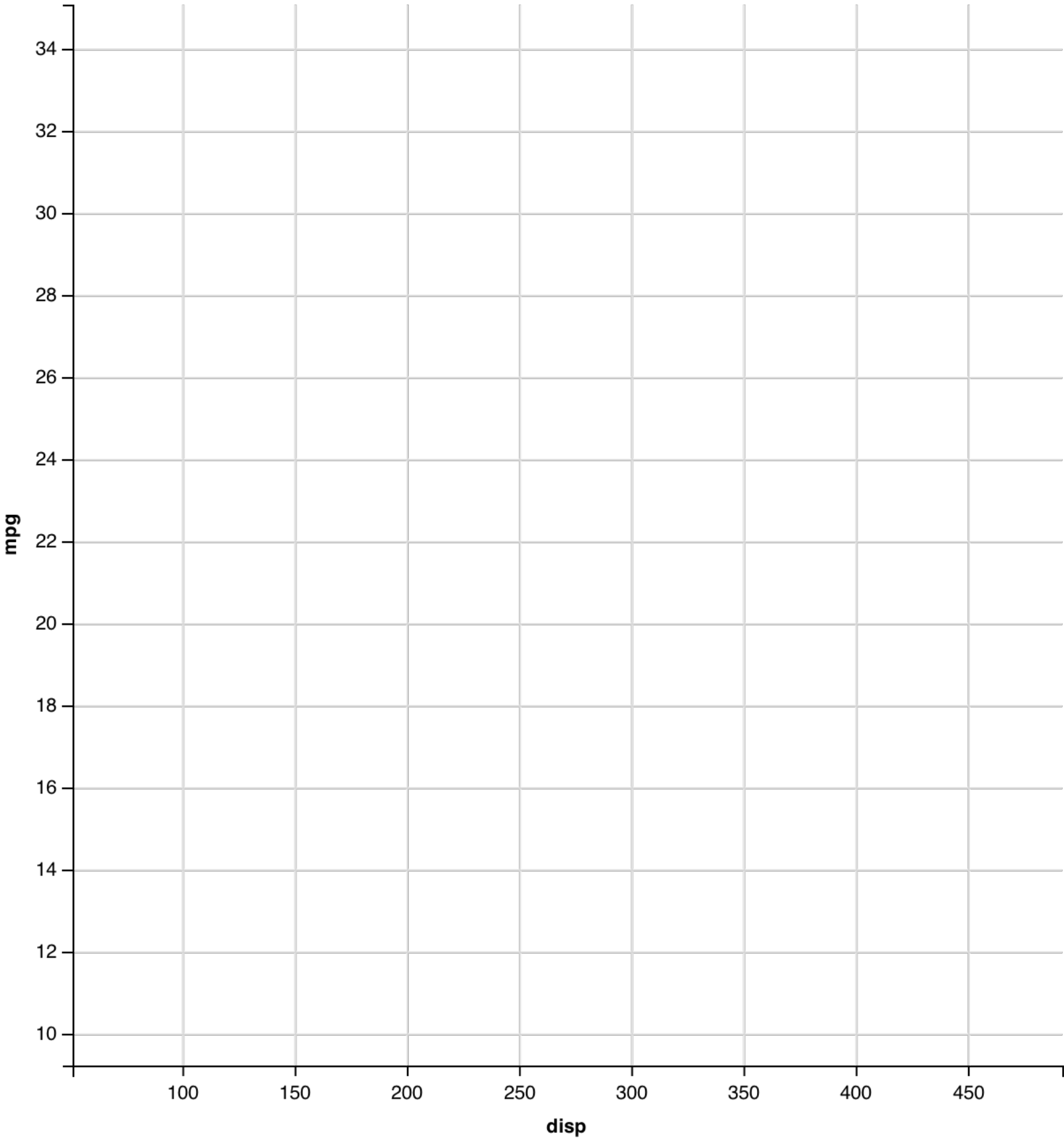
mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data



mpg	cyl	displacement	horsepower
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-
-



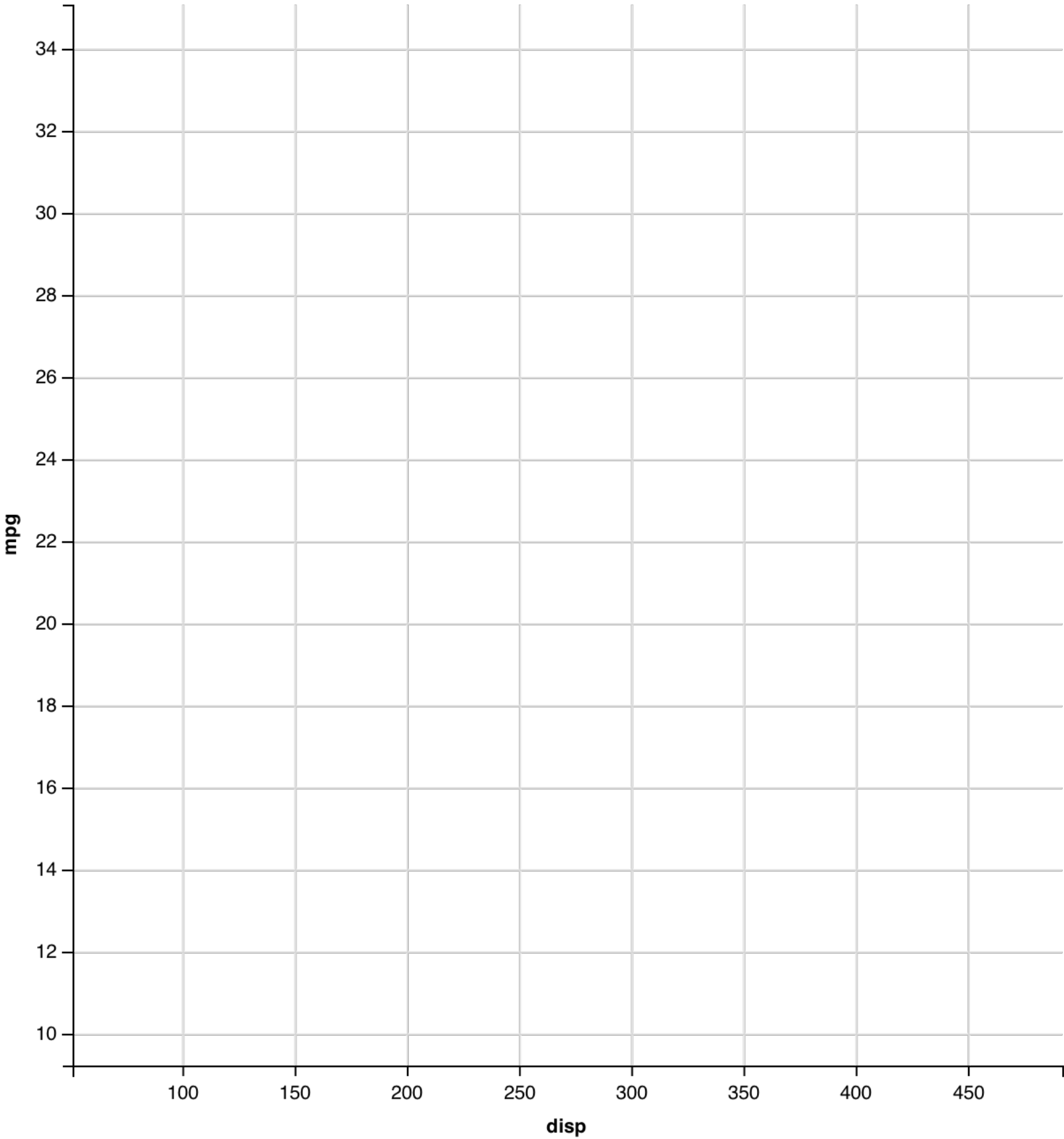
data geom



mpg	cyl	disp	hp	
21.0	6	160.0	2	●
21.0	6	160.0	2	●
22.8	4	108.0	1	●
21.4	6	258.0	2	●
18.7	8	360.0	3	●
18.1	6	225.0	2	●
14.3	8	360.0	5	●
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	●
17.8	6	167.6	2	●
16.4	8	275.8	3	●
17.3	8	275.8	3	●
15.2	8	275.8	3	●
10.4	8	472.0	4	●
10.4	8	460.0	4	●
14.7	8	440.0	4	●
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●

data

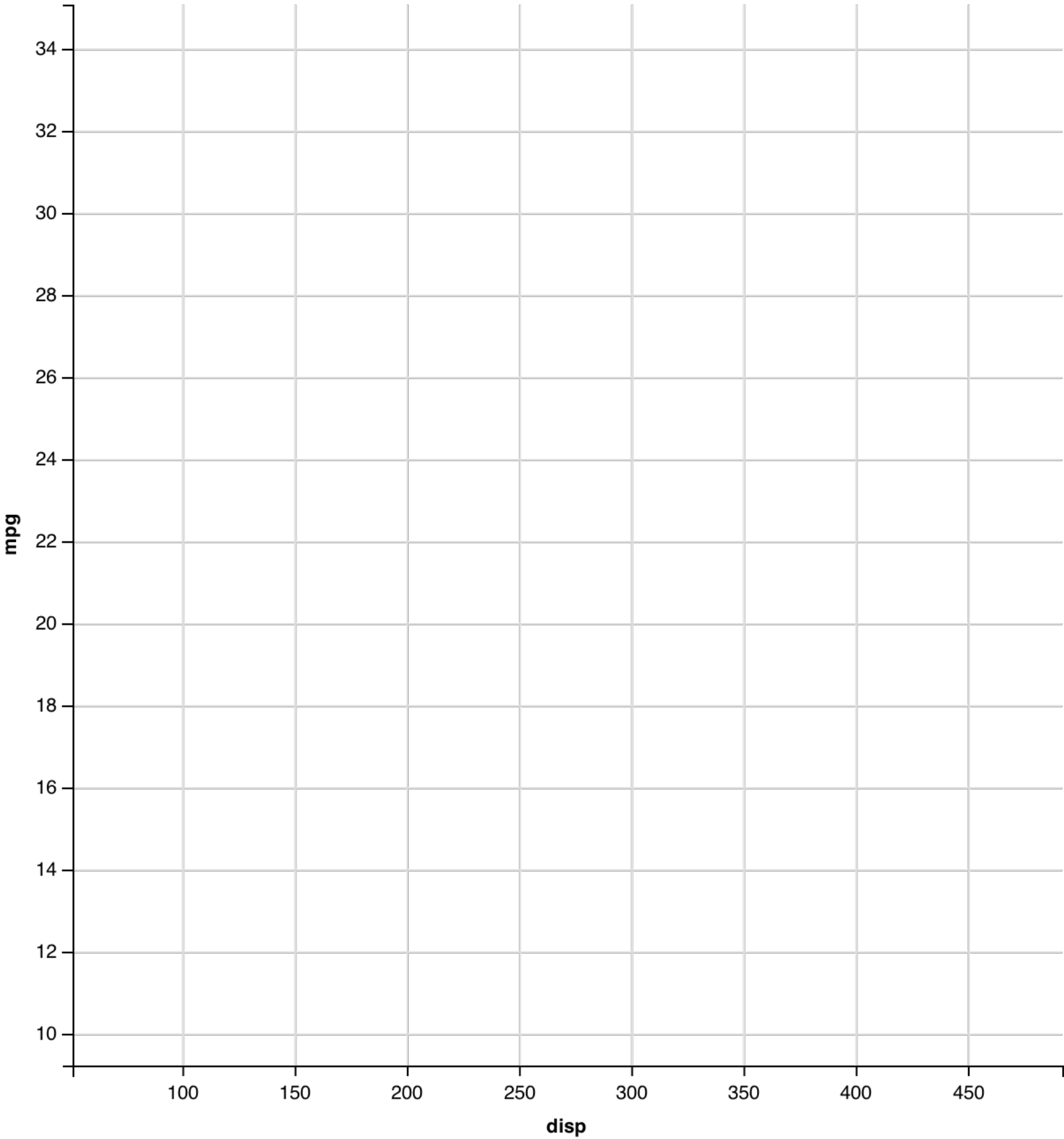
geom



mappings

fill
↑↓

mpg	cyl	displacement	horsepower
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

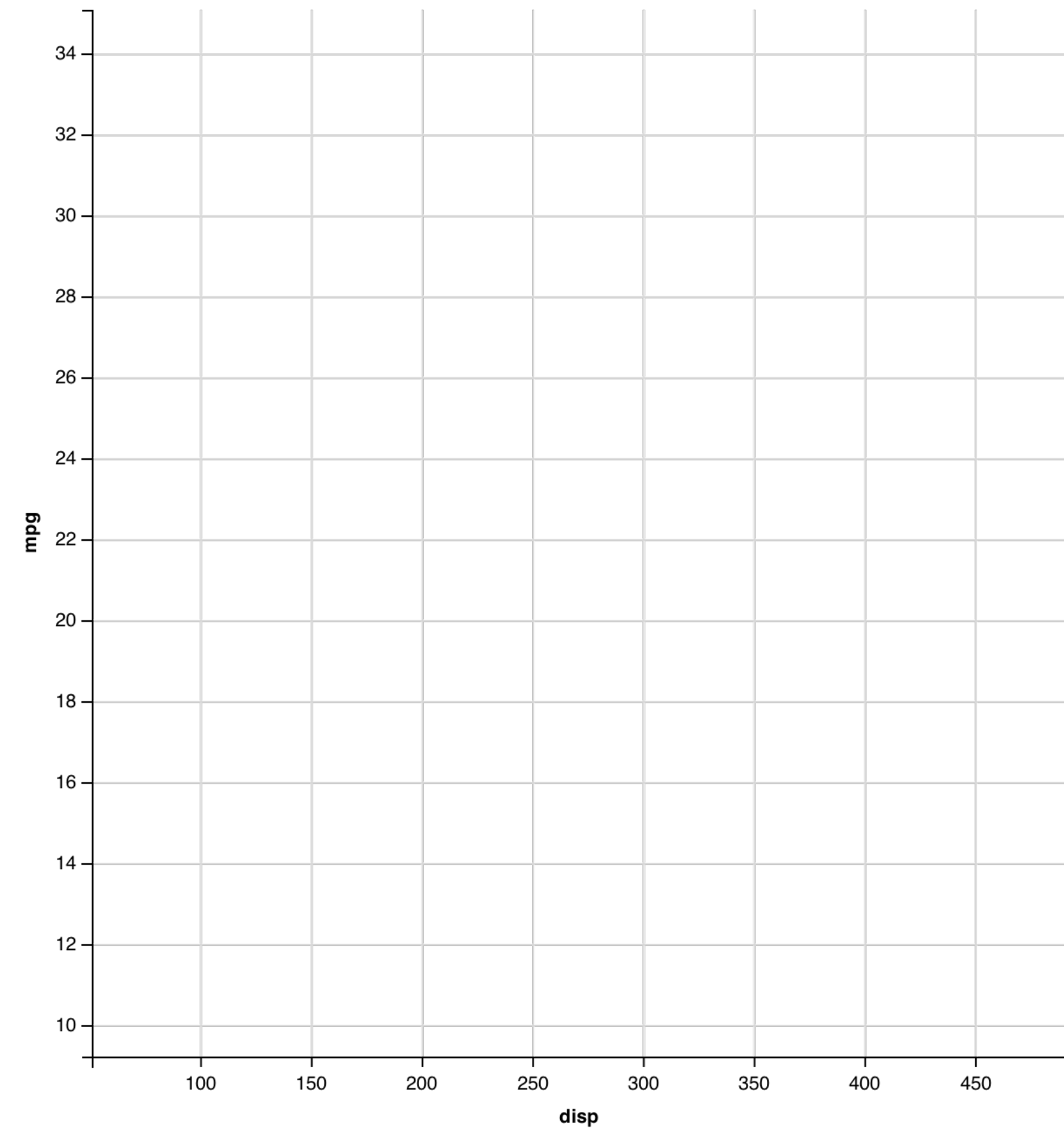


data geom



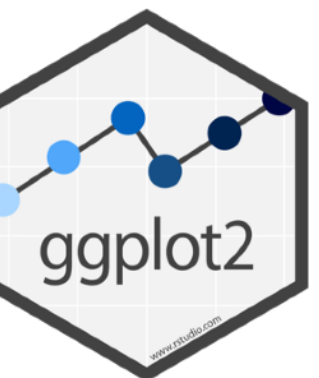
mappings

mpg	cyl	disp	hp	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●



data

geom

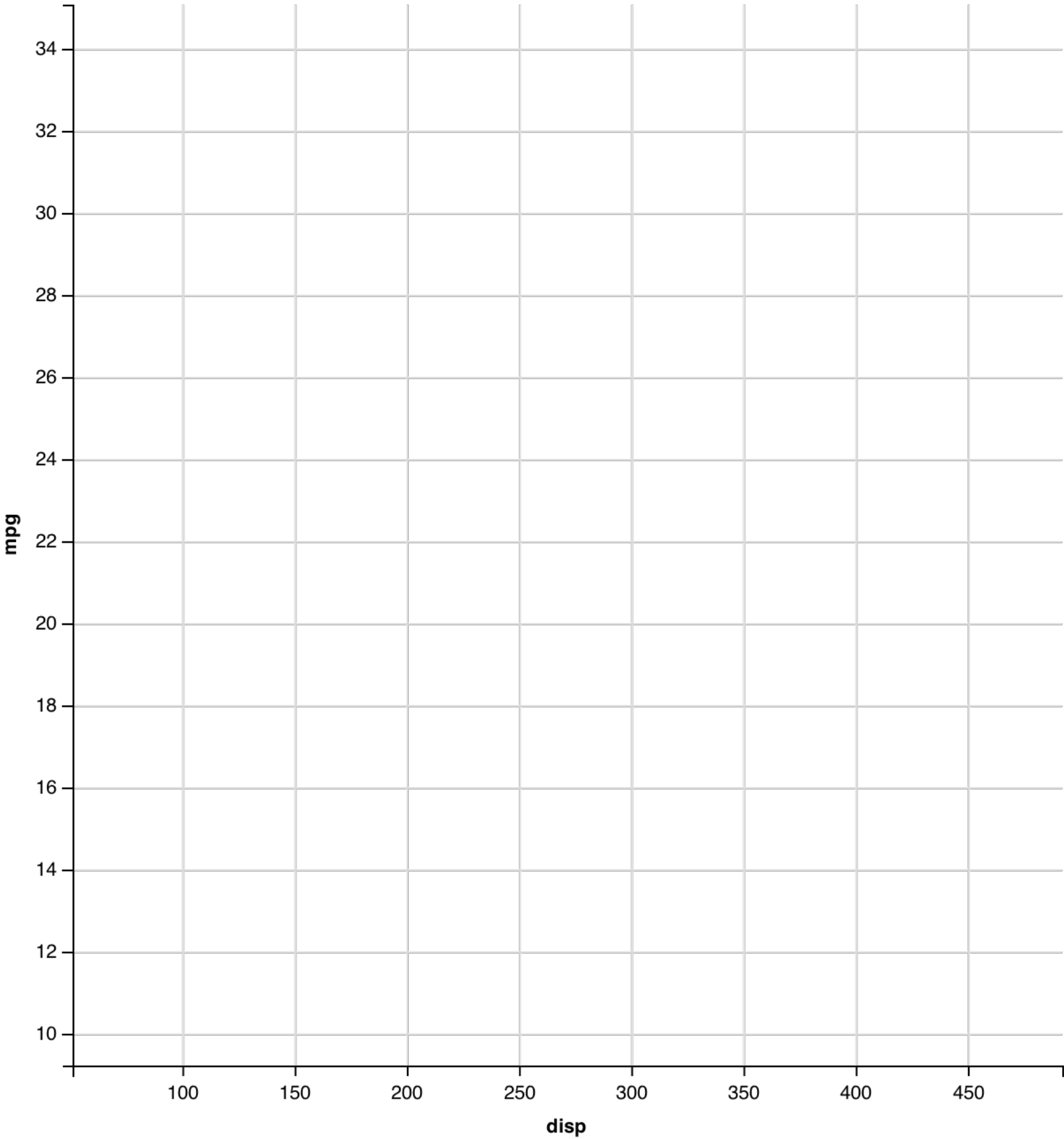


mappings

shape

fill

mpg	cyl	displacement	horsepower	
21.0	6 +	160.0	2	+
21.0	6 +	160.0	2	+
22.8	4 ●	108.0	1	●
21.4	6 +	258.0	2	+
18.7	8 ◆	360.0	3	◆
18.1	6 +	225.0	2	+
14.3	8 ◆	360.0	5	◆
24.4	4 ●	146.7	1	●
22.8	4 ●	140.8	1	●
19.2	6 +	167.6	2	+
17.8	6 +	167.6	2	+
16.4	8 ◆	275.8	3	◆
17.3	8 ◆	275.8	3	◆
15.2	8 ◆	275.8	3	◆
10.4	8 ◆	472.0	4	◆
10.4	8 ◆	460.0	4	◆
14.7	8 ◆	440.0	4	◆
32.4	4 ●	78.7	1	●
30.4	4 ●	75.7	1	●
33.9	4 ●	71.1	1	●



data

geom

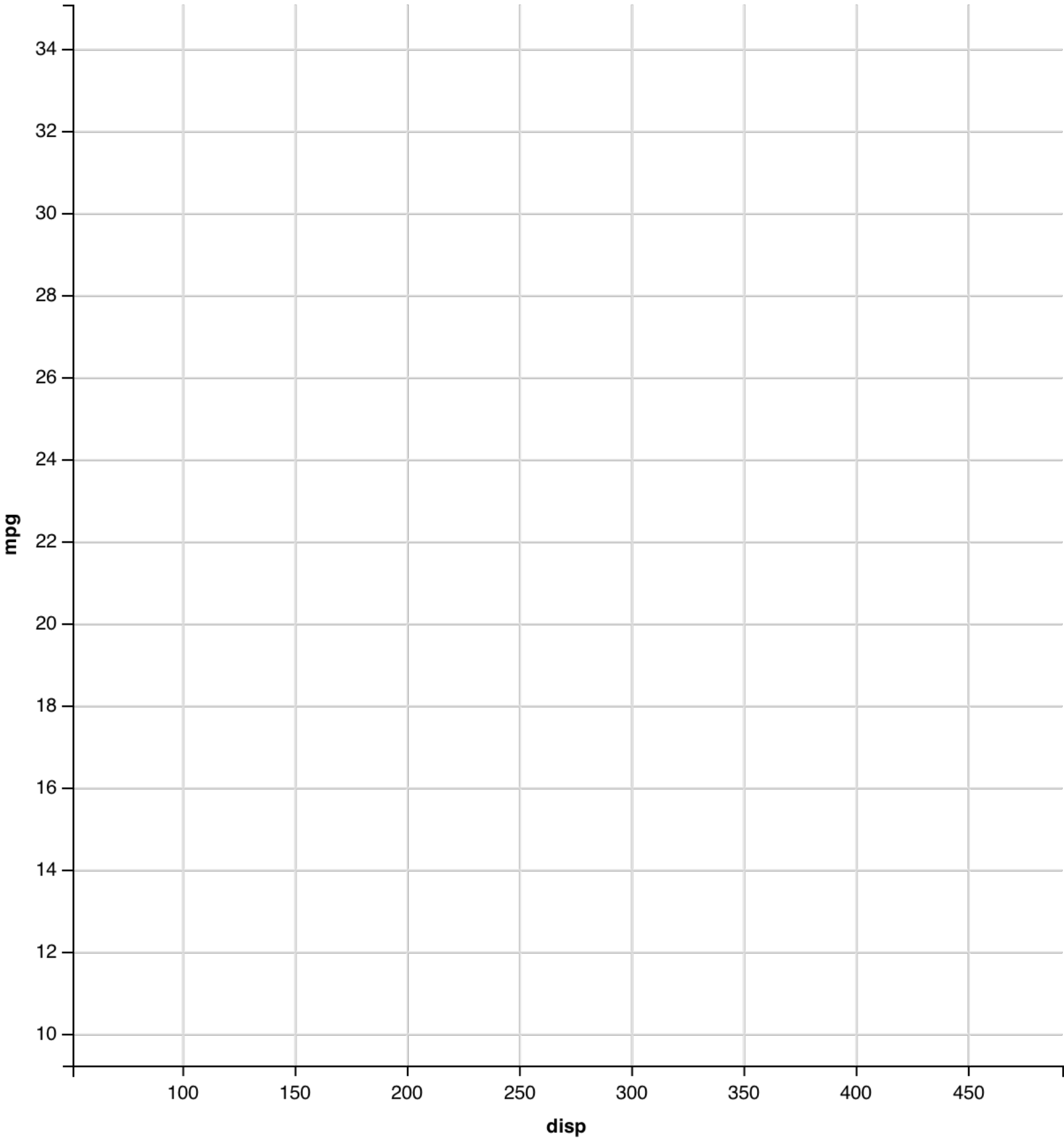


mappings

shape

fill

mpg	cyl	displacement	horsepower	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●



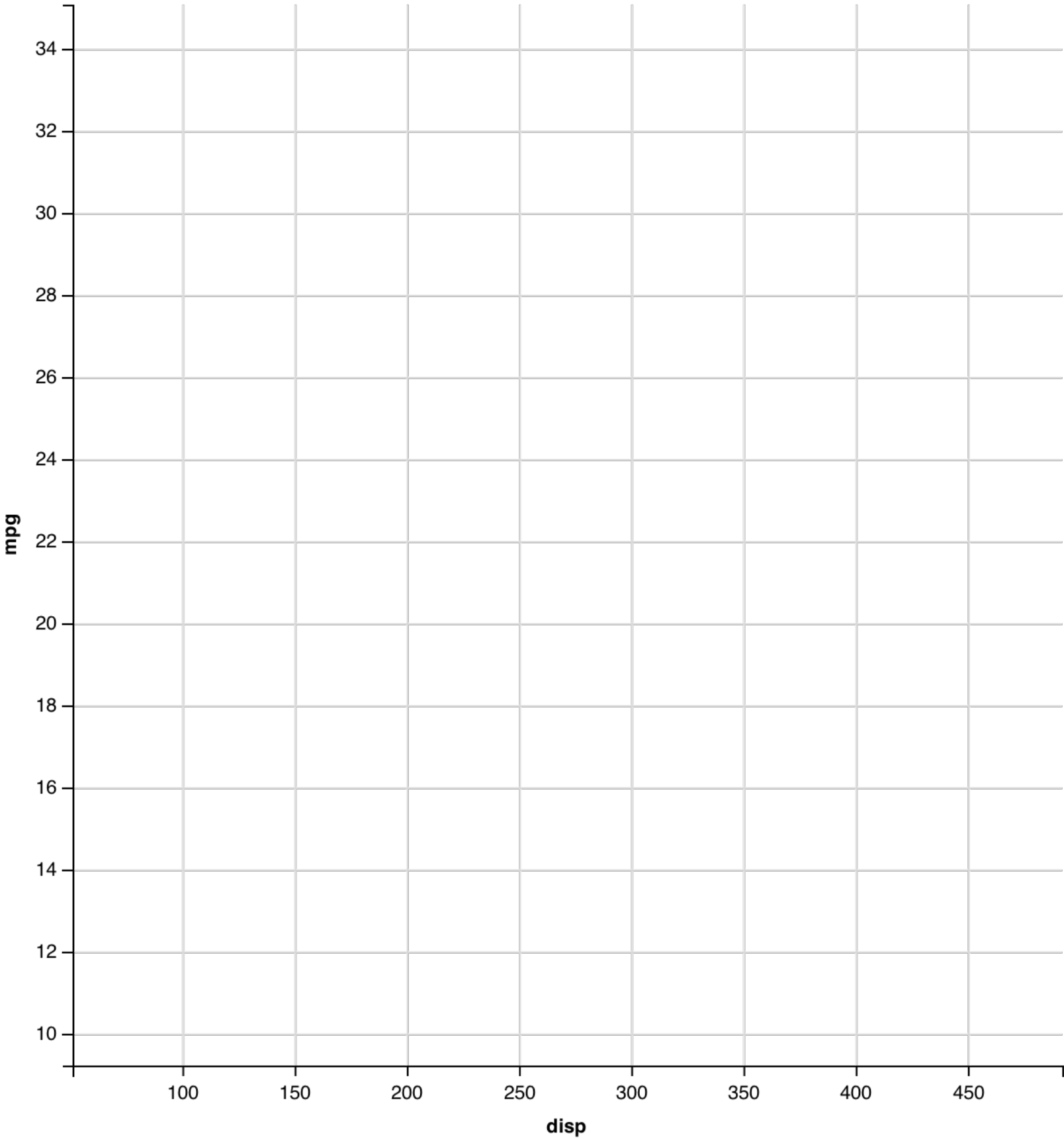
data

geom



mappings

	shape	x	fill	
	↕	↕	↕	
mpg	cyl	displacement	hp	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●



data geom



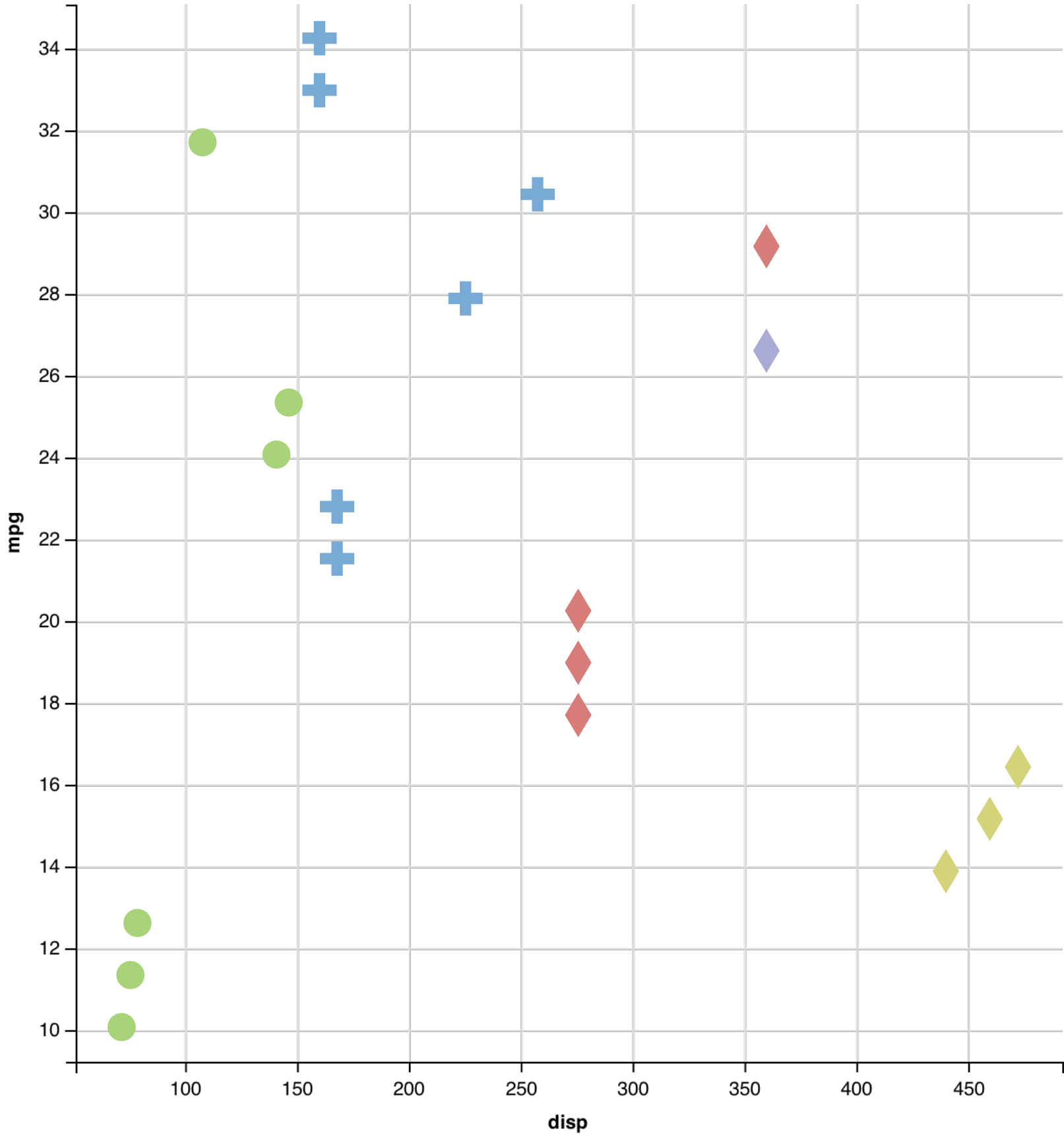
mappings

shape

x

fill

mpg	cyl	dispx	hp	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●

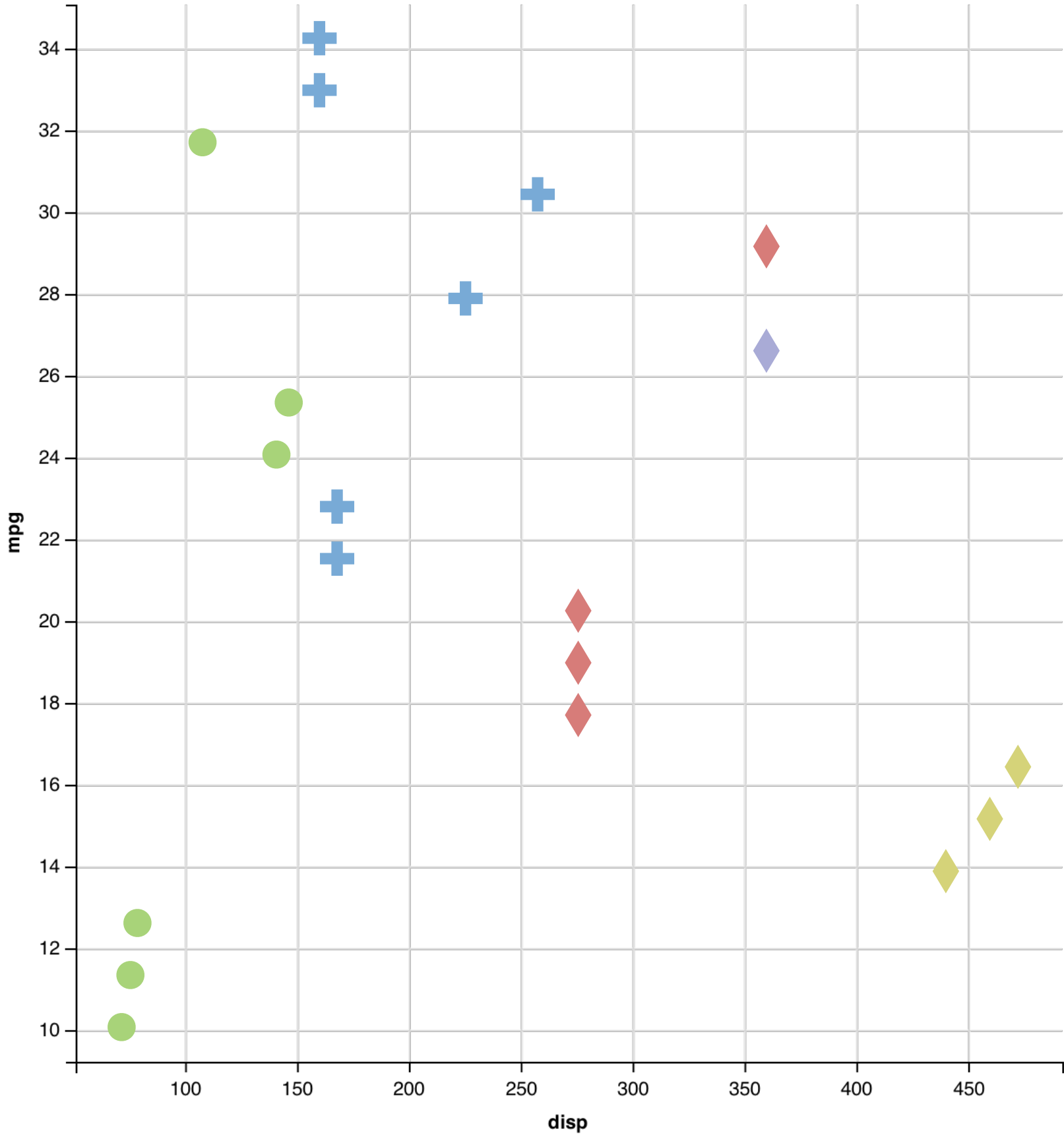


data geom



mappings

mpg	shape	x	fill	
	cyl	displacement	hp	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●

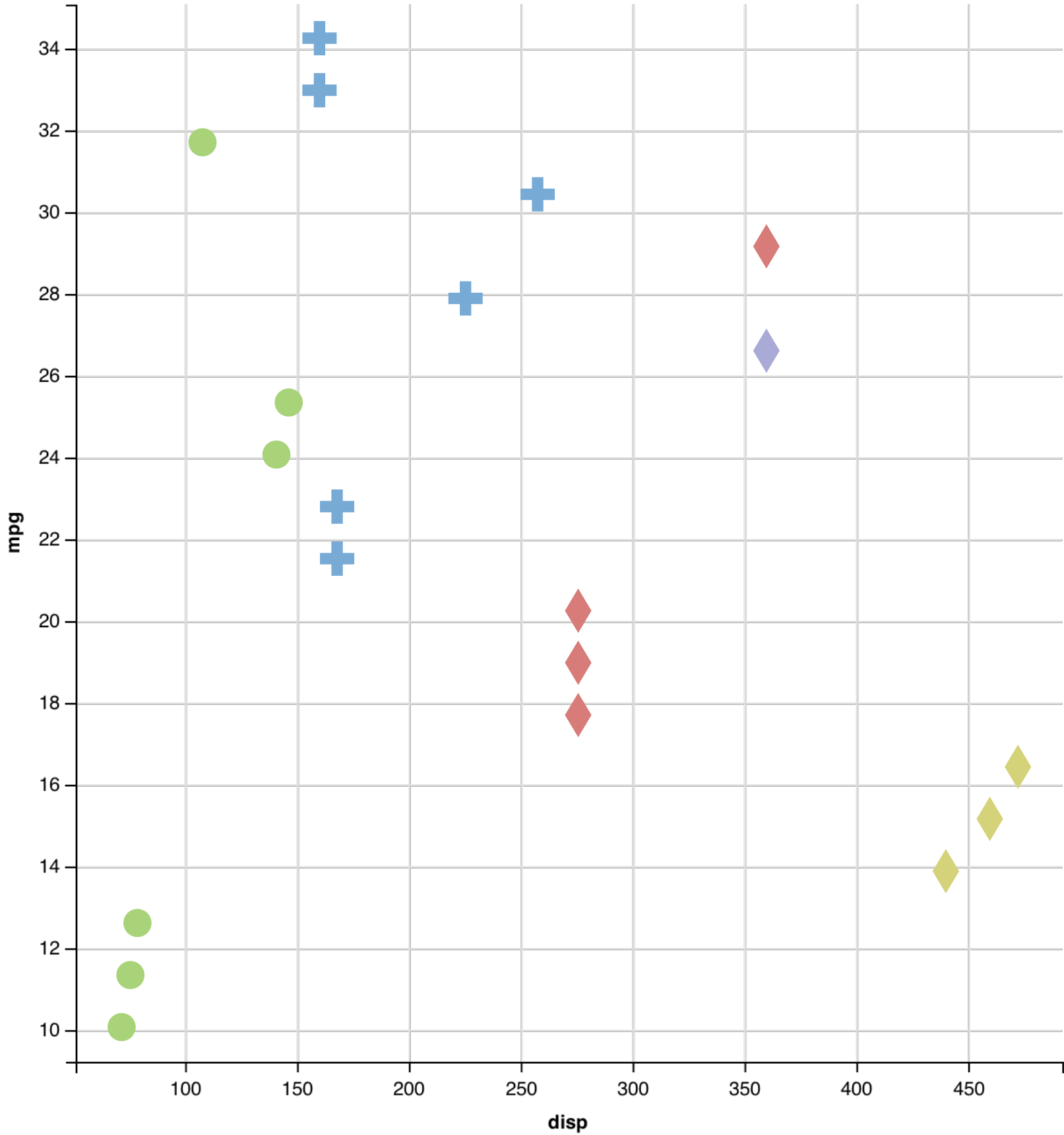


data geom



mappings

y	shape	x	fill	
↕	↕	↕	↕	
mpg	cyl	displacement	hp	
21.0	6	160.0	2	+
21.0	6	160.0	2	+
22.8	4	108.0	1	●
21.4	6	258.0	2	+
18.7	8	360.0	3	◆
18.1	6	225.0	2	+
14.3	8	360.0	5	◆
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	+
17.8	6	167.6	2	+
16.4	8	275.8	3	◆
17.3	8	275.8	3	◆
15.2	8	275.8	3	◆
10.4	8	472.0	4	◆
10.4	8	460.0	4	◆
14.7	8	440.0	4	◆
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●

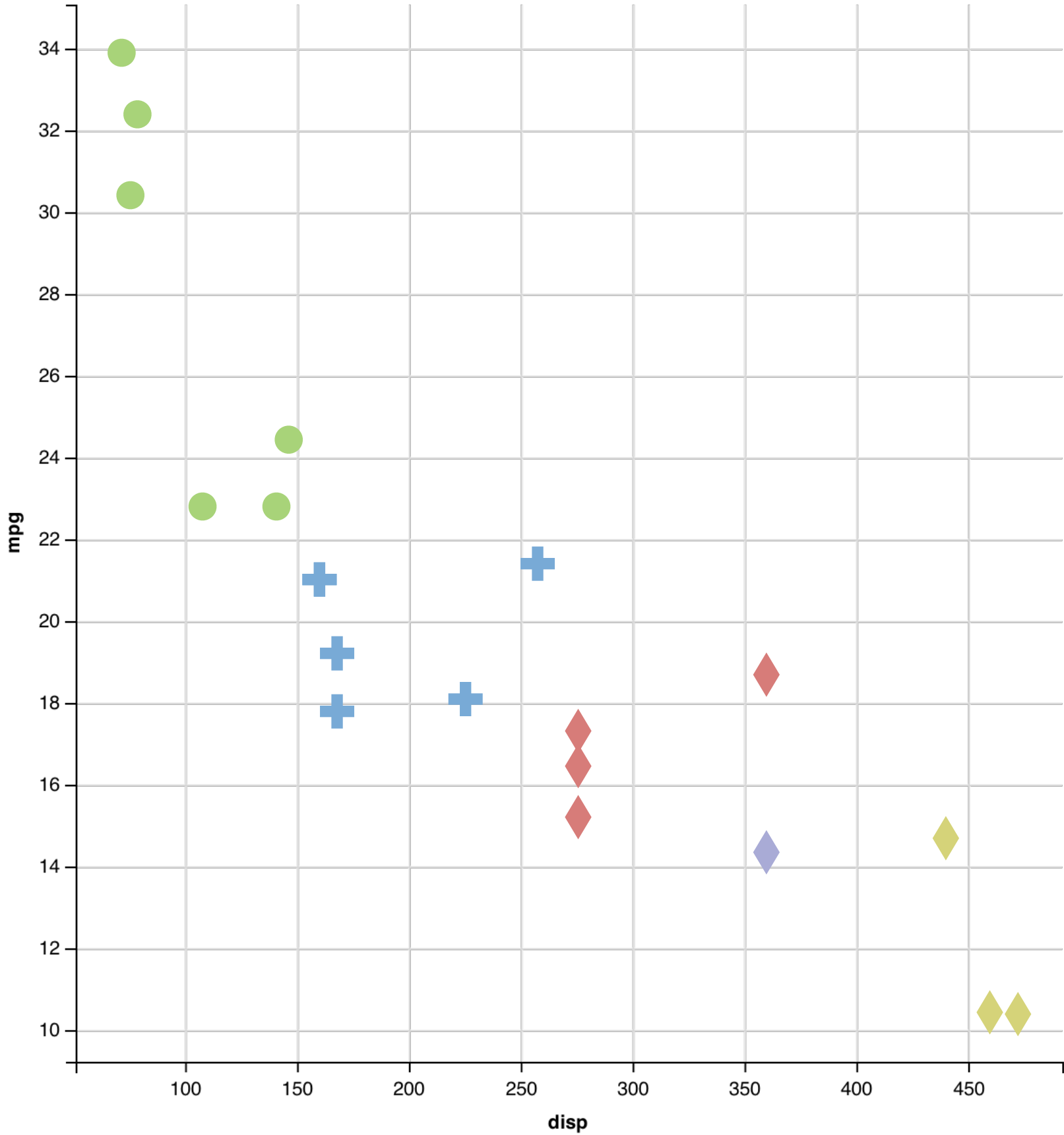


data geom



mappings

y	shape	x	fill
mpg	cyl	displacement	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1



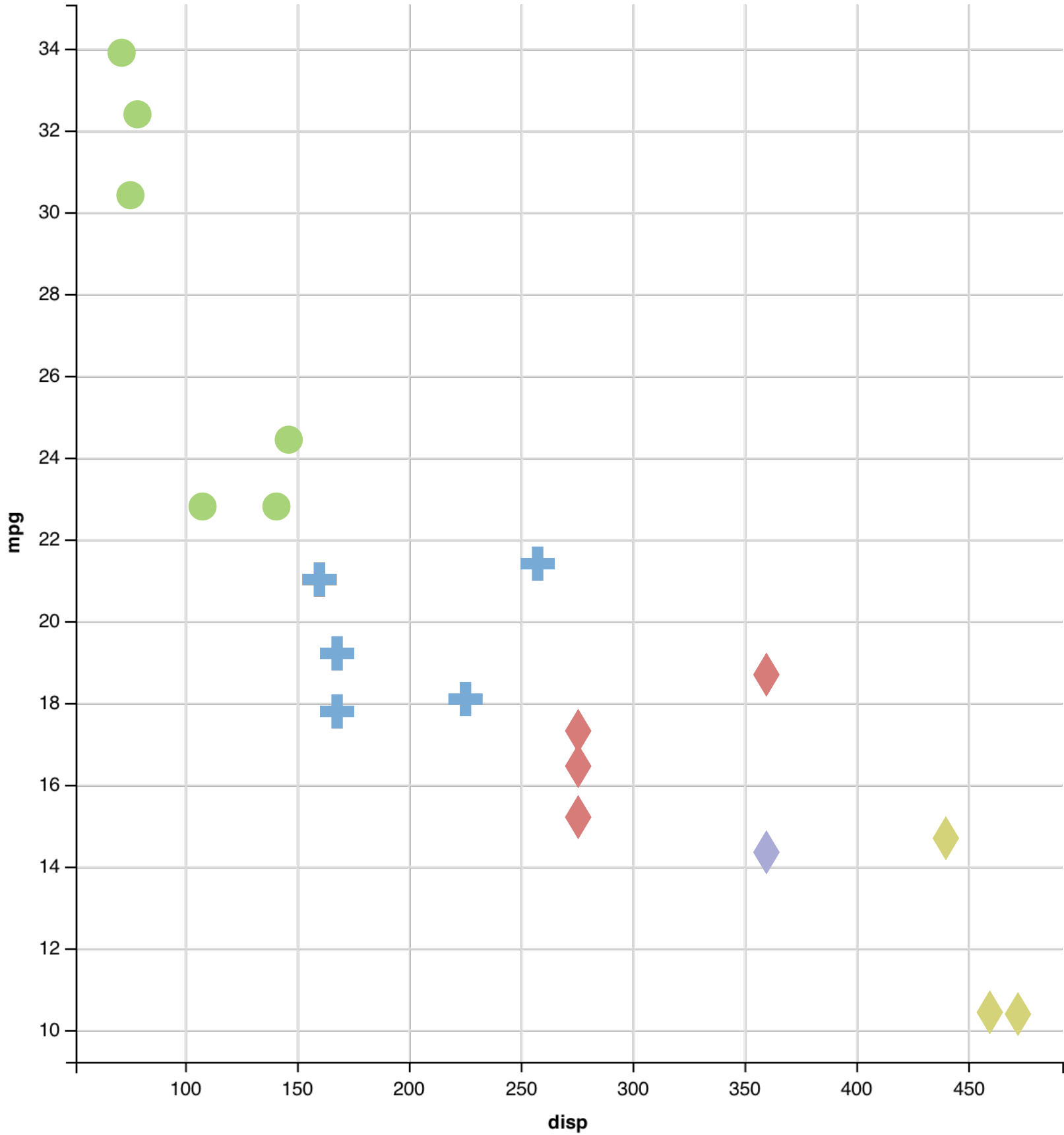
data

geom



mappings

y	shape	x	fill
mpg	cyl	displacement	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

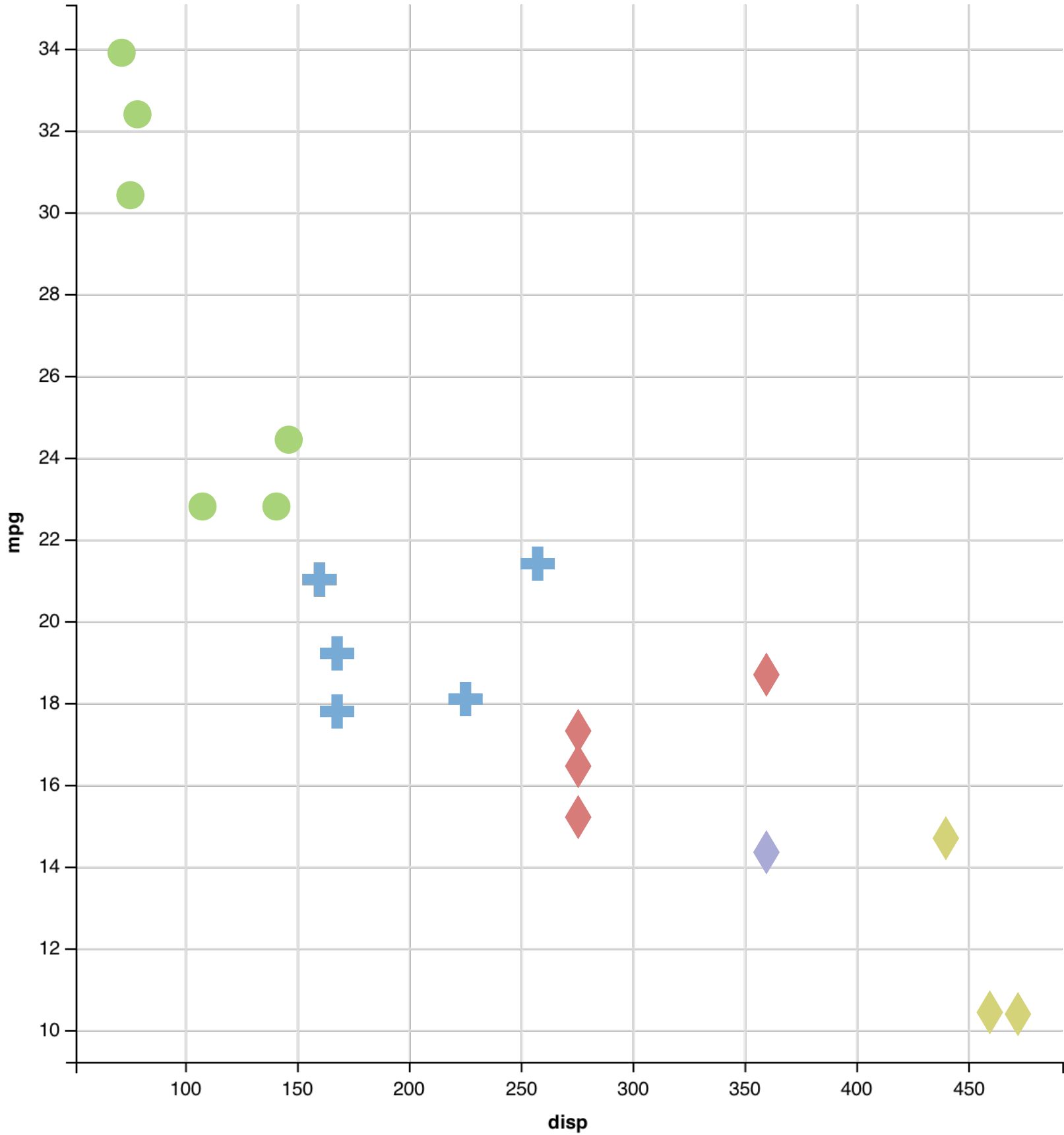


data geom



mappings

y		x		
↑↓		↑↓		
mpg	cyl	displacement	horsepower	
21.0	6	160.0	2	point
21.0	6	160.0	2	point
22.8	4	108.0	1	point
21.4	6	258.0	2	point
18.7	8	360.0	3	point
18.1	6	225.0	2	point
14.3	8	360.0	5	point
24.4	4	146.7	1	point
22.8	4	140.8	1	point
19.2	6	167.6	2	point
17.8	6	167.6	2	point
16.4	8	275.8	3	point
17.3	8	275.8	3	point
15.2	8	275.8	3	point
10.4	8	472.0	4	point
10.4	8	460.0	4	point
14.7	8	440.0	4	point
32.4	4	78.7	1	point
30.4	4	75.7	1	point
33.9	4	71.1	1	point



data

geom
points
lines

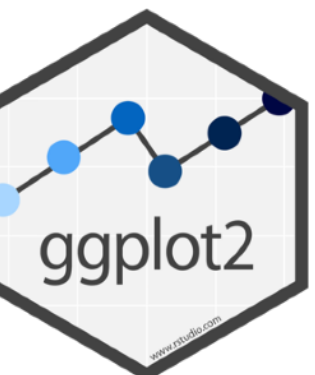
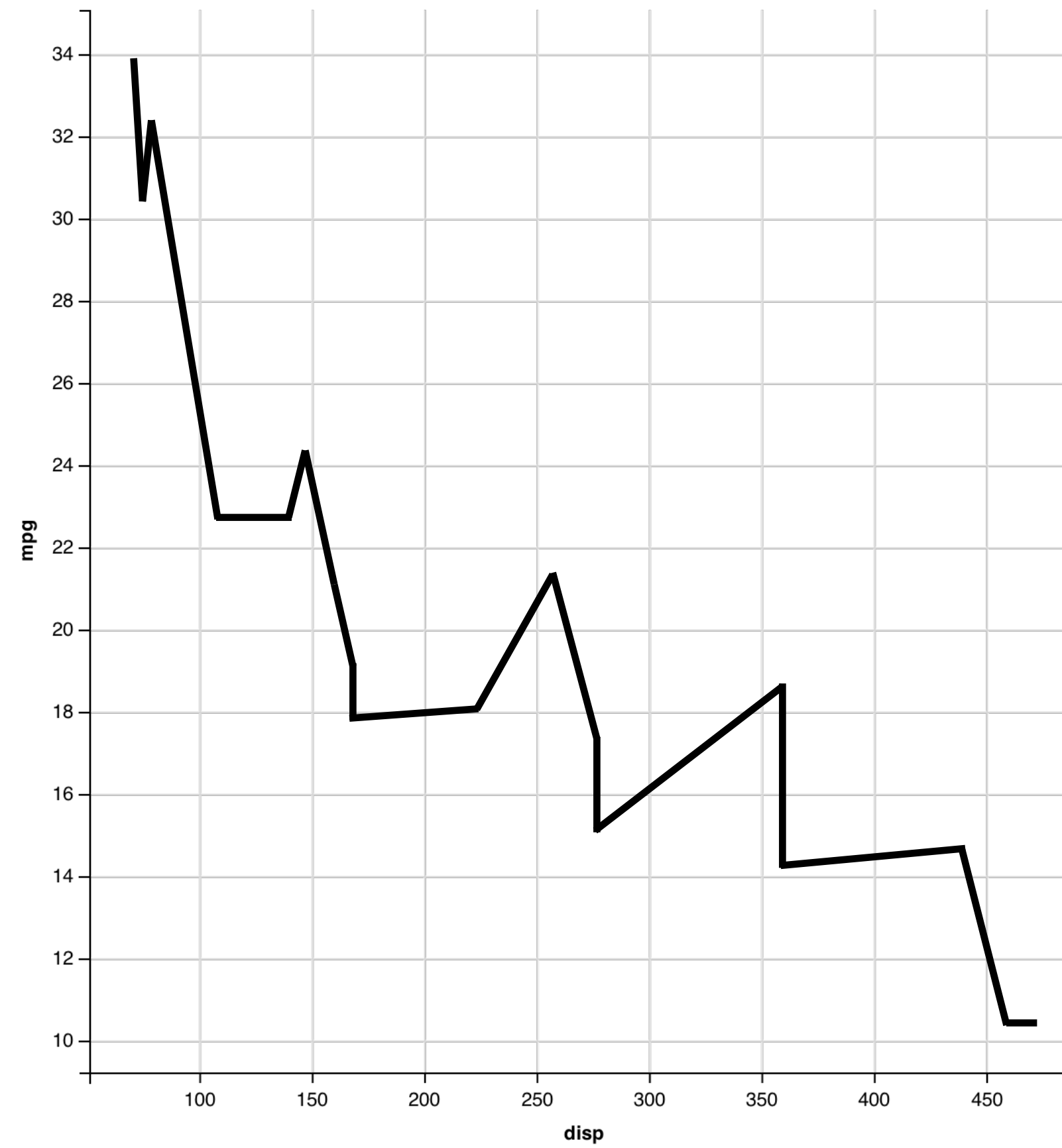


mappings

y		x		
↕		↕		
mpg	cyl	disp	hp	
21.0	6	160.0	2	
21.0	6	160.0	2	
22.8	4	108.0	1	
21.4	6	258.0	2	
18.7	8	360.0	3	
18.1	6	225.0	2	
14.3	8	360.0	5	
24.4	4	146.7	1	
22.8	4	140.8	1	
19.2	6	167.6	2	
17.8	6	167.6	2	
16.4	8	275.8	3	
17.3	8	275.8	3	
15.2	8	275.8	3	
10.4	8	472.0	4	
10.4	8	460.0	4	
14.7	8	440.0	4	
32.4	4	78.7	1	
30.4	4	75.7	1	
33.9	4	71.1	1	

data

geom
points
lines



mappings

y

↑

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

x

↑

—

—

—

—

—

—

—

—

—

—

—

—

—

—

—

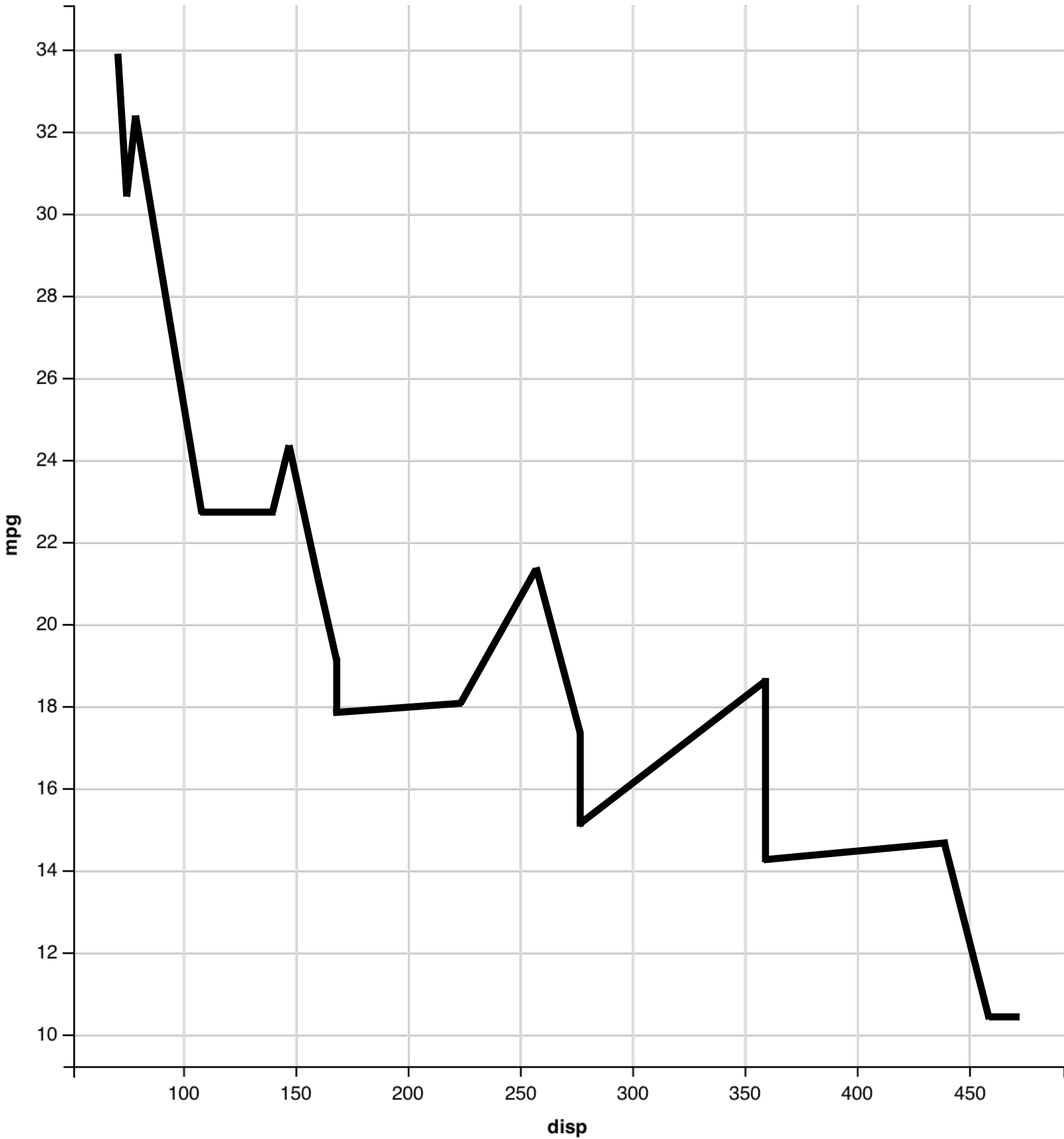
—

—

—

—

—



data

geom
points
lines

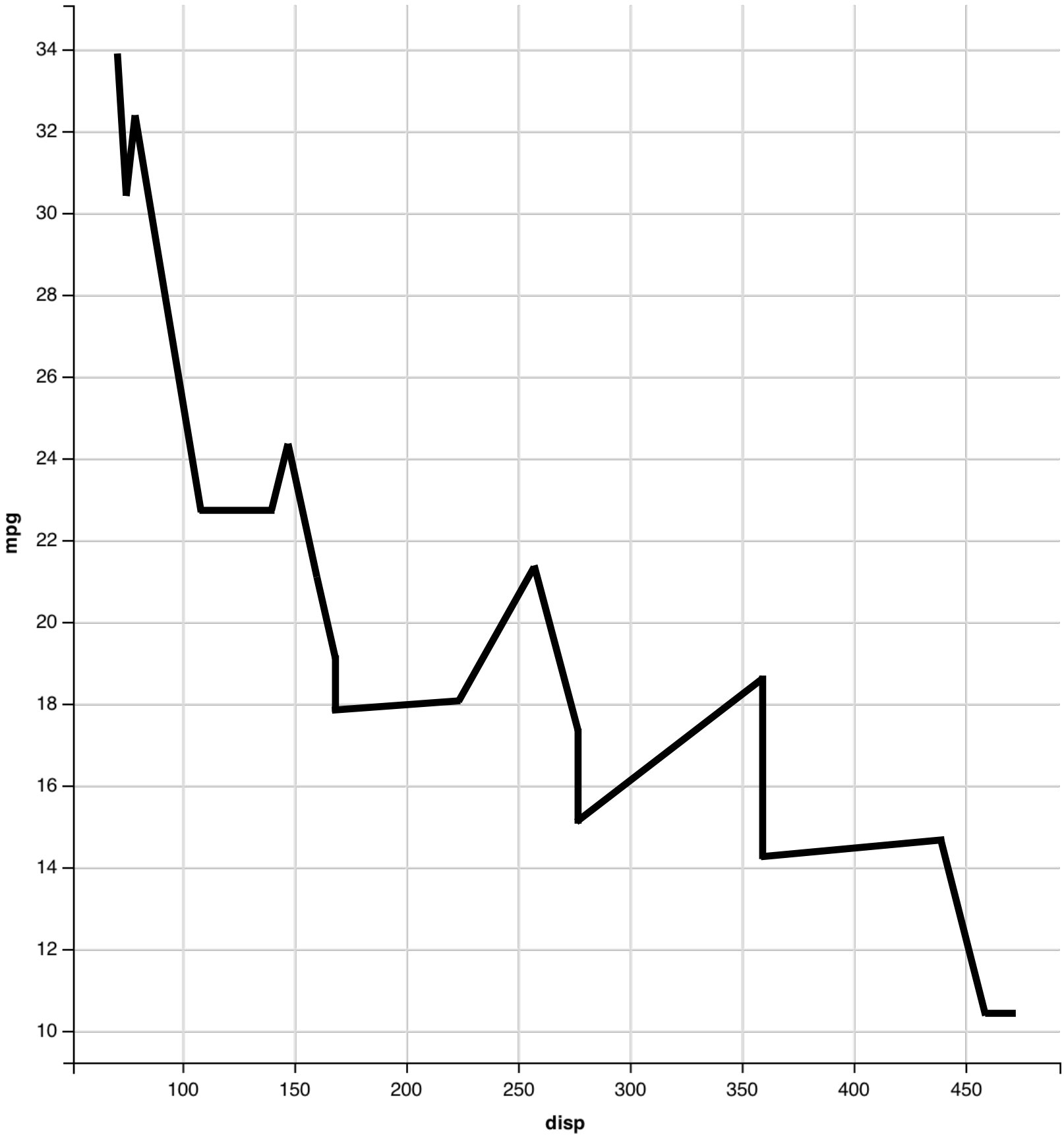


mappings

y

↑

mpg	cyl	dispx	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1



data

- geom
- points
 - lines
 - bars

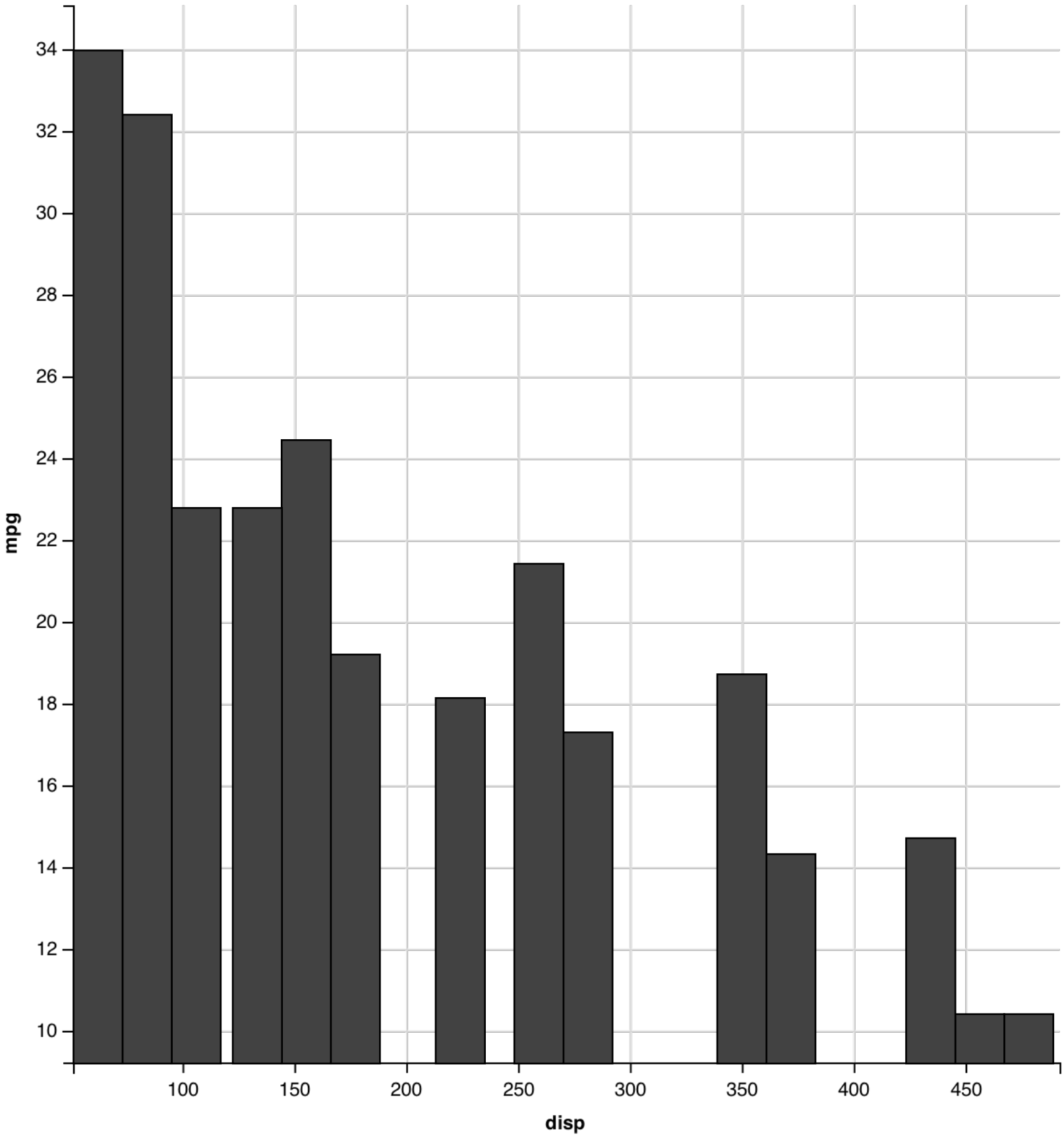


mappings

y

↑

mpg	cyl	dispx	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1



data

- geom
- points
 - lines
 - bars



mappings

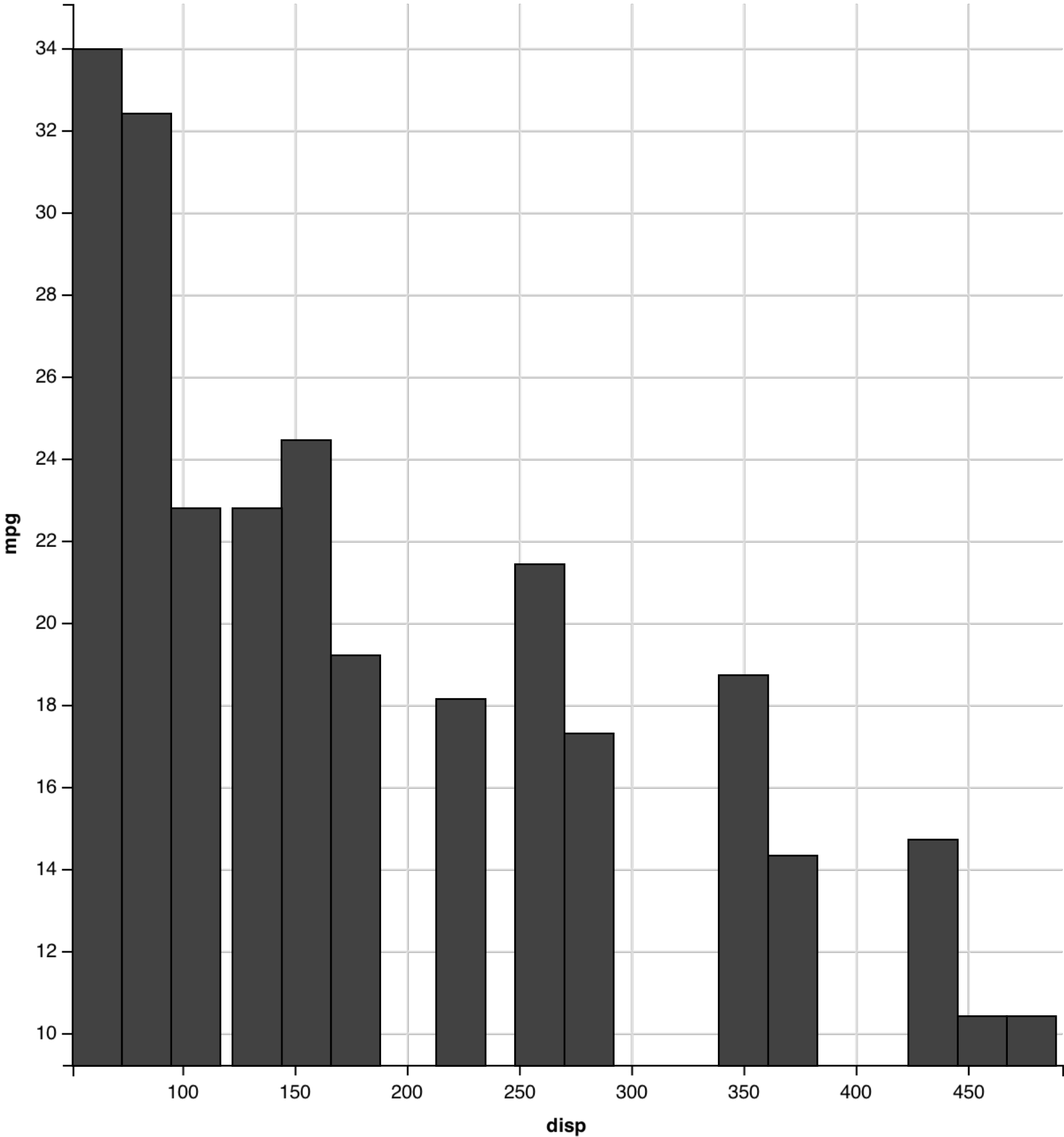
y

↑↓

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

x

↑↓

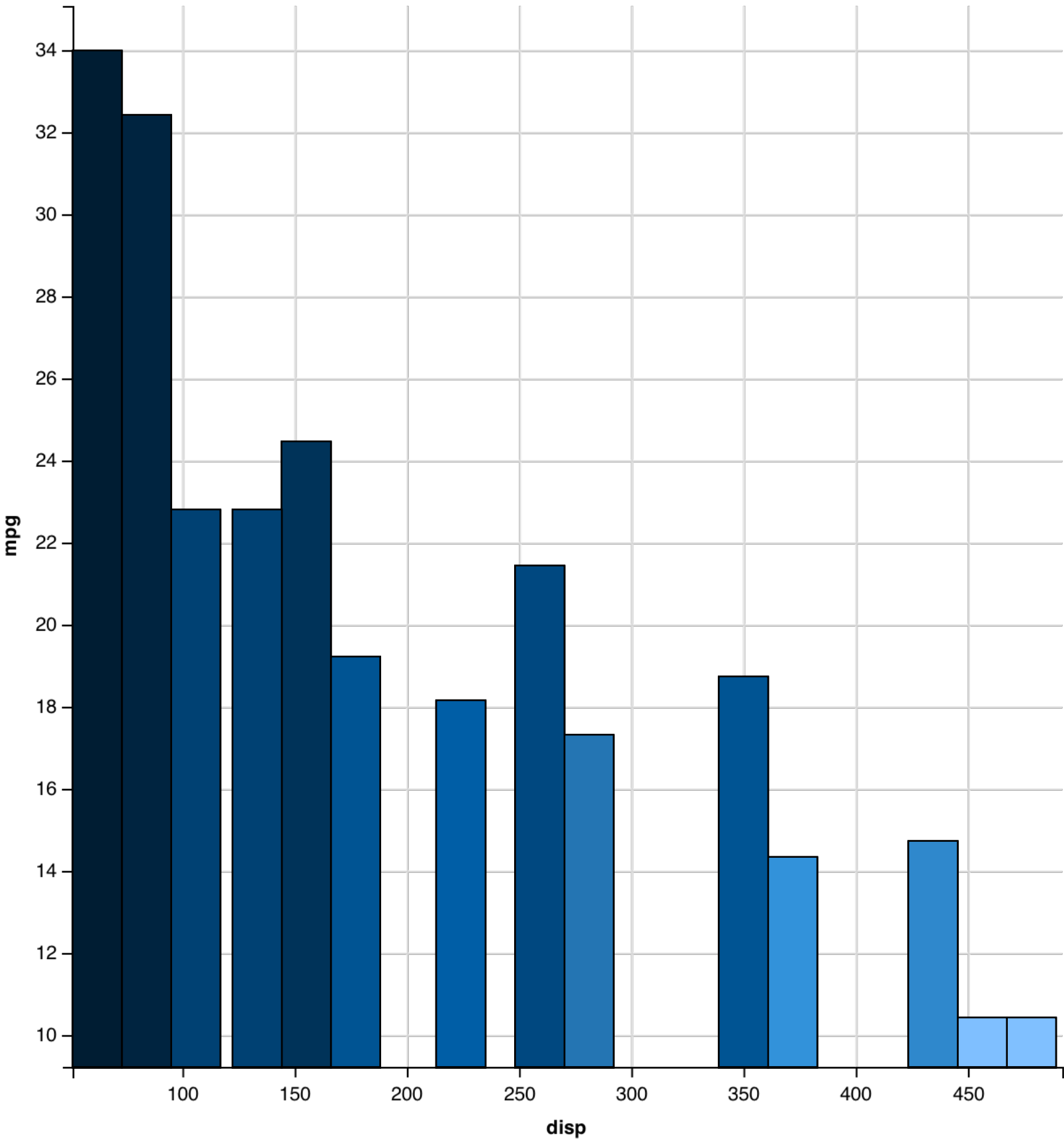
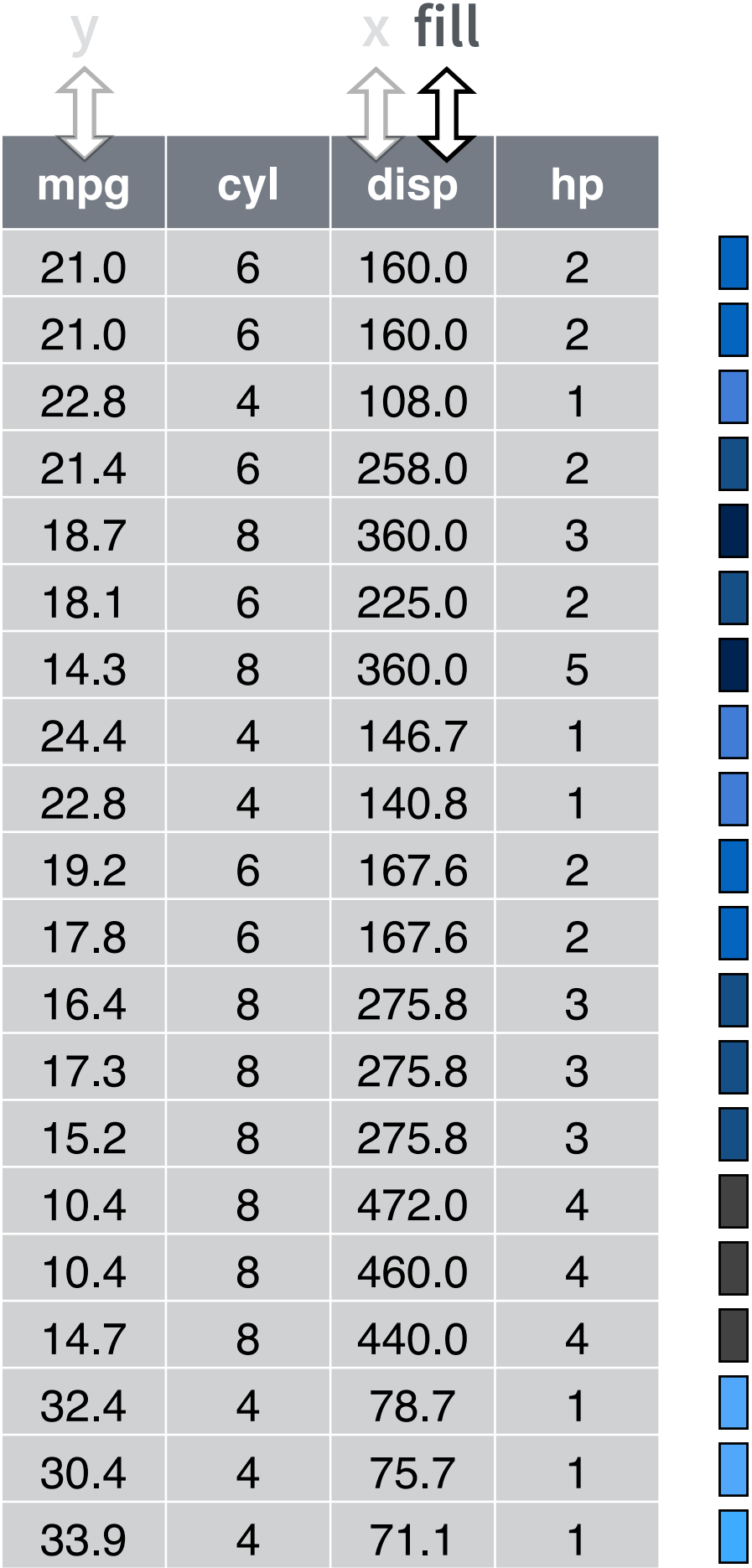


data

- geom
- points
 - lines
 - bars



mappings

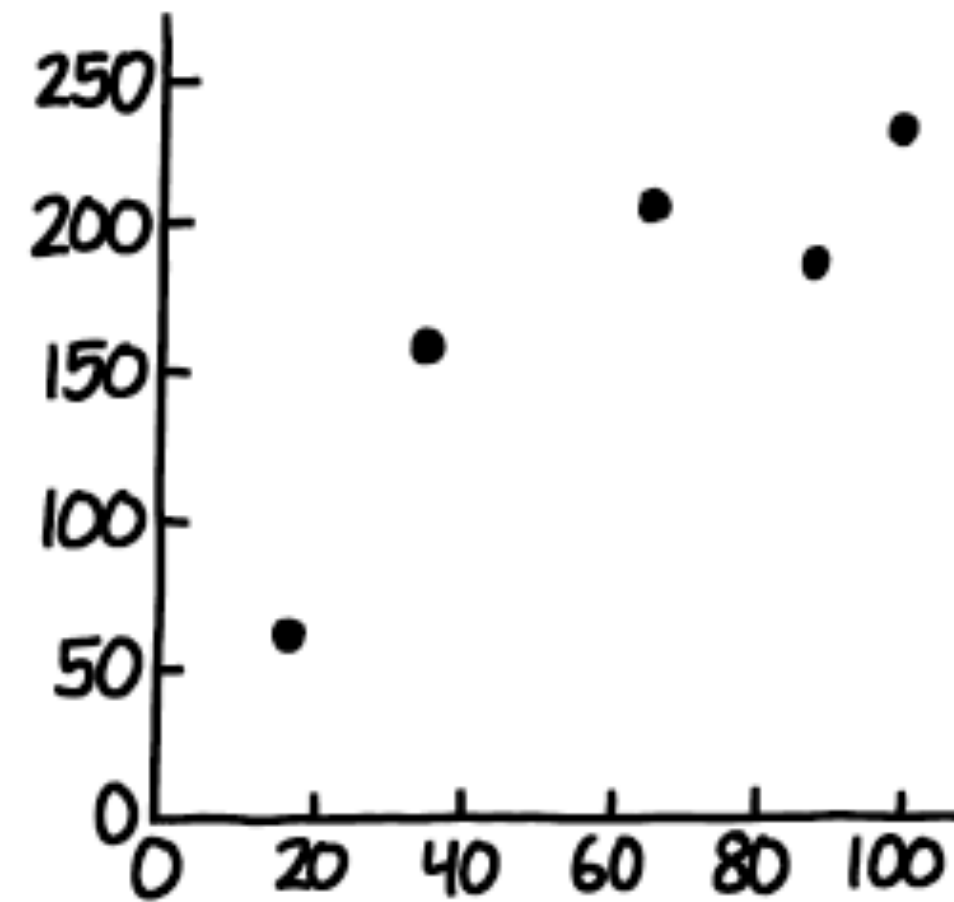


data

- geom
- points
 - lines
 - bars



VARIABLE 1: X AXIS
VARIABLE 2: Y AXIS



VARIABLE 1: HUE
VARIABLE 2: LABEL

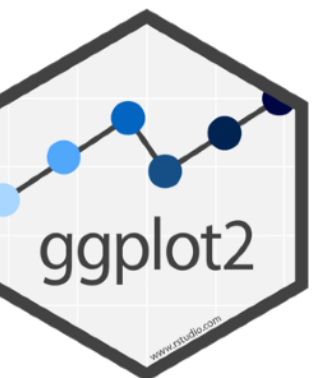


DESIGN TIP: YOU CAN MAKE YOUR GRAPHS MORE SPACE-EFFICIENT BY USING HUE AND LABEL FOR THE FIRST TWO VARIABLES, INSTEAD OF ONLY TURNING TO THEM ONCE YOU'VE USED UP THE X AND Y AXES.

To make a graph

[template]

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```



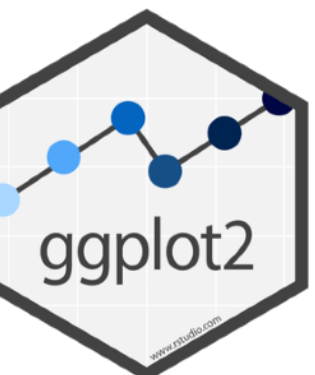
To make a graph

1. Pick a **data** set

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

data

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```



To make a graph

mpg	cyl	disp	hp
21.0	6	160.0	2
21.0	6	160.0	2
22.8	4	108.0	1
21.4	6	258.0	2
18.7	8	360.0	3
18.1	6	225.0	2
14.3	8	360.0	5
24.4	4	146.7	1
22.8	4	140.8	1
19.2	6	167.6	2
17.8	6	167.6	2
16.4	8	275.8	3
17.3	8	275.8	3
15.2	8	275.8	3
10.4	8	472.0	4
10.4	8	460.0	4
14.7	8	440.0	4
32.4	4	78.7	1
30.4	4	75.7	1
33.9	4	71.1	1

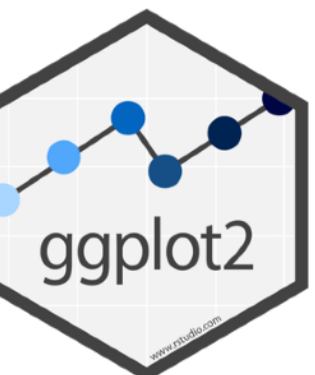
data

geom

1. Pick a **data** set

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

2. Choose a **geom**
to display cases



To make a graph

mappings

mpg	cyl	disp	hp	fill
21.0	6	160.0	2	●
21.0	6	160.0	2	●
22.8	4	108.0	1	●
21.4	6	258.0	2	●
18.7	8	360.0	3	●
18.1	6	225.0	2	●
14.3	8	360.0	5	●
24.4	4	146.7	1	●
22.8	4	140.8	1	●
19.2	6	167.6	2	●
17.8	6	167.6	2	●
16.4	8	275.8	3	●
17.3	8	275.8	3	●
15.2	8	275.8	3	●
10.4	8	472.0	4	●
10.4	8	460.0	4	●
14.7	8	440.0	4	●
32.4	4	78.7	1	●
30.4	4	75.7	1	●
33.9	4	71.1	1	●

data

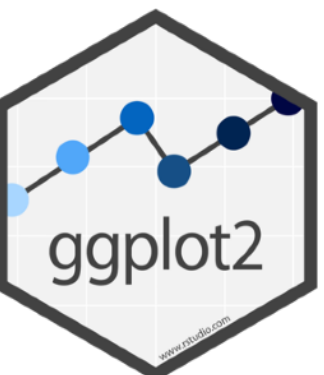
geom

1. Pick a **data** set

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

2. Choose a **geom**
to display cases

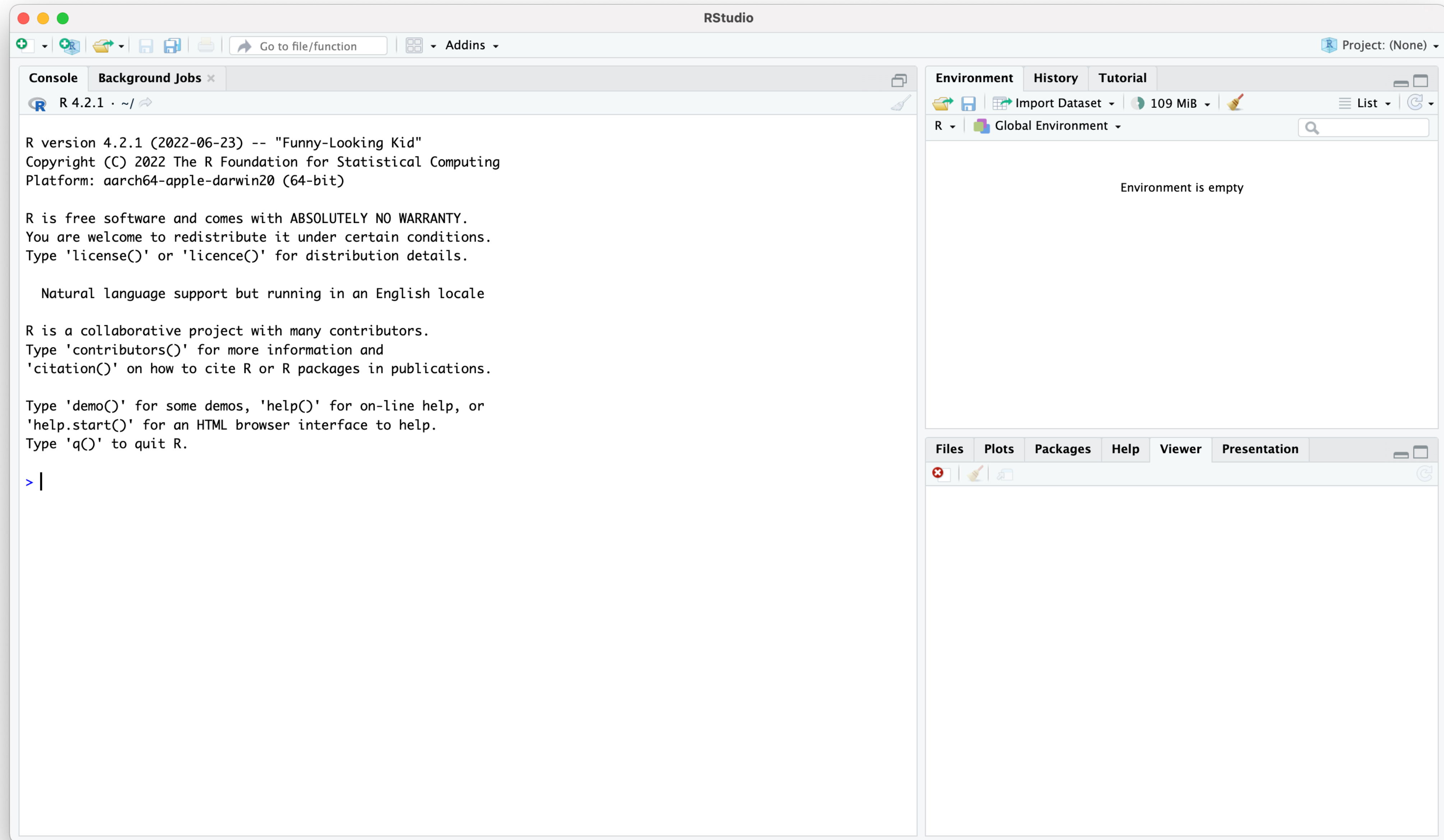
3. **Map** aesthetic
properties to
variables



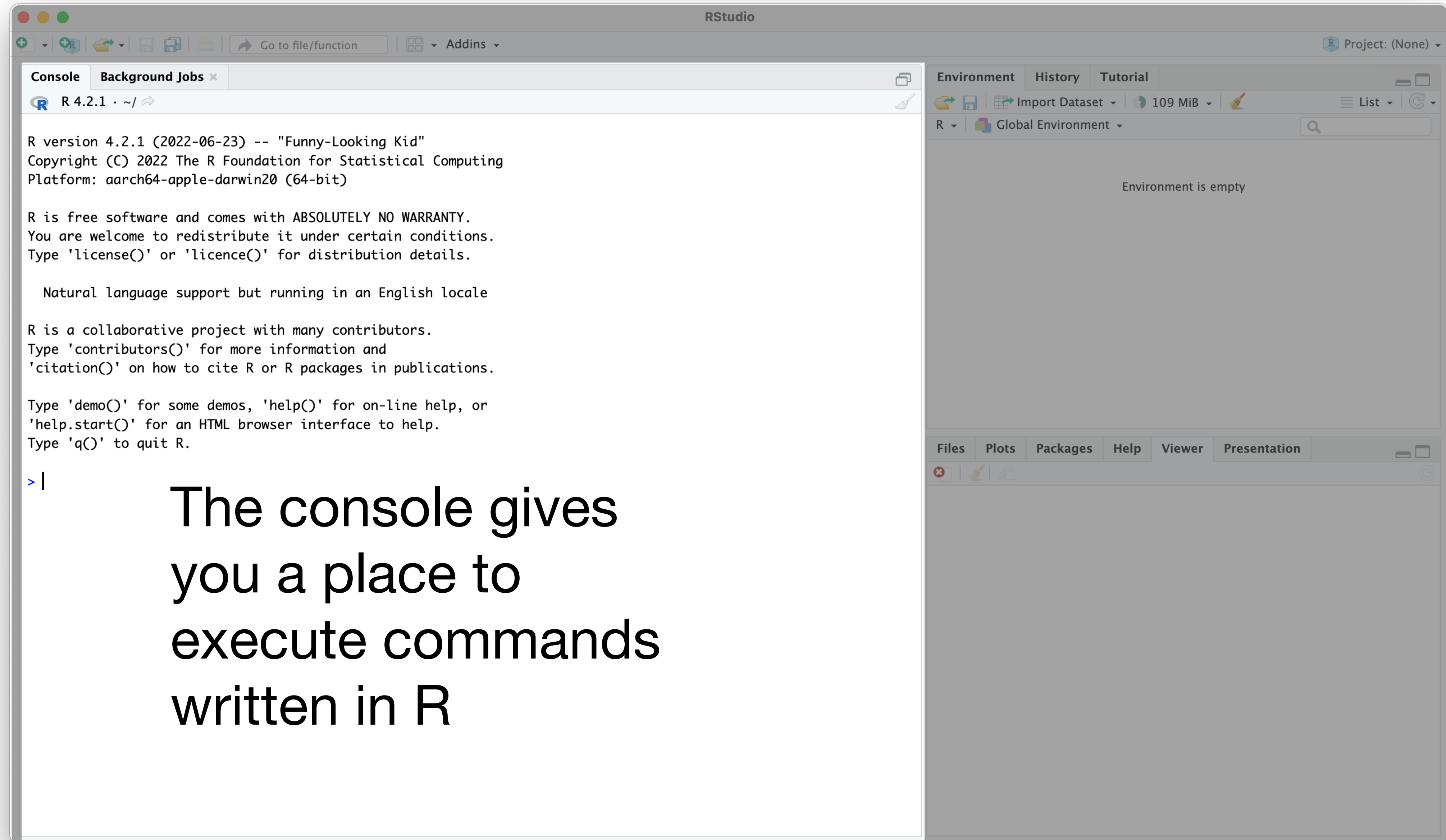
RStudio
(let's start!)



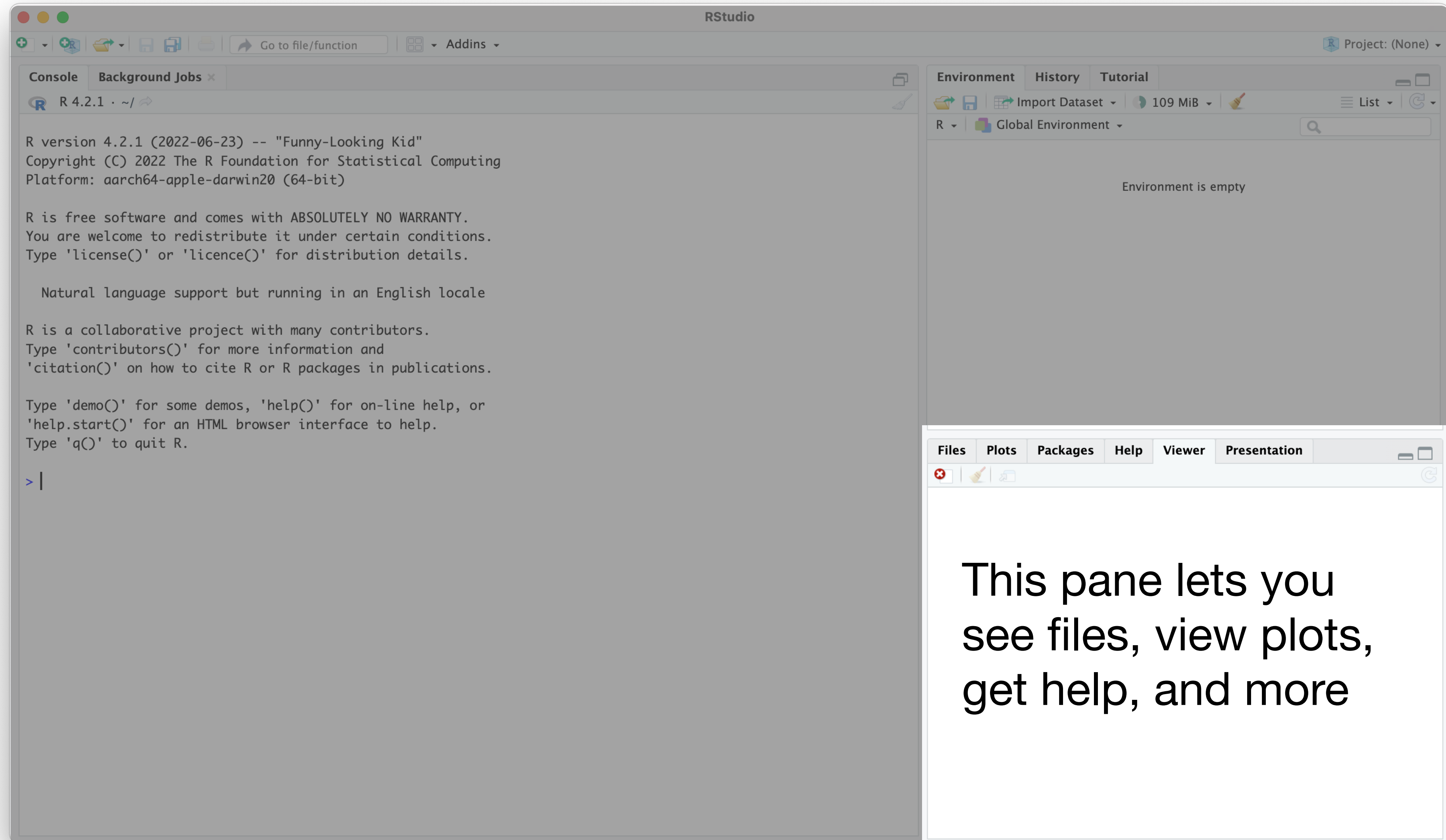
RStudio



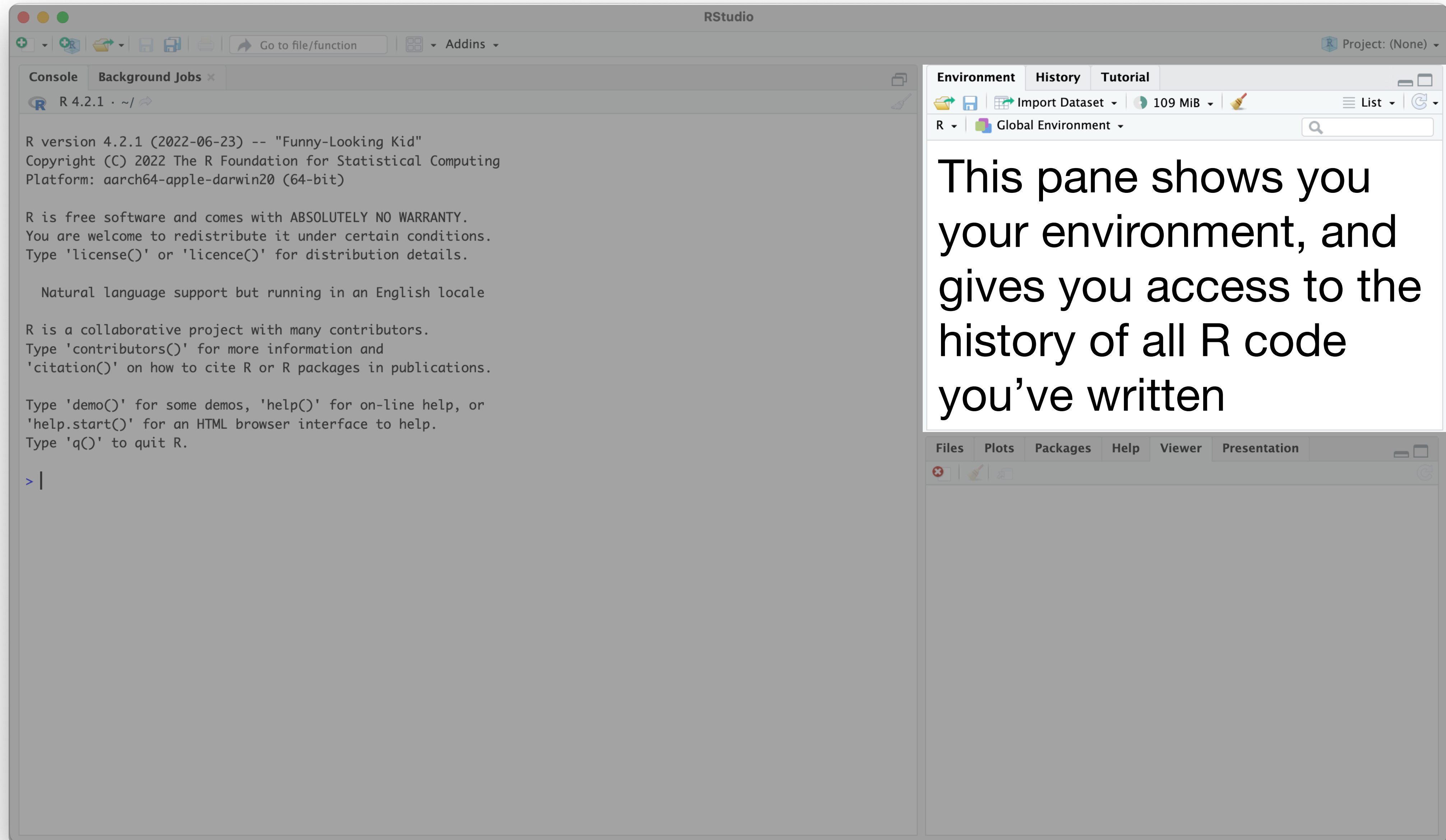
RStudio



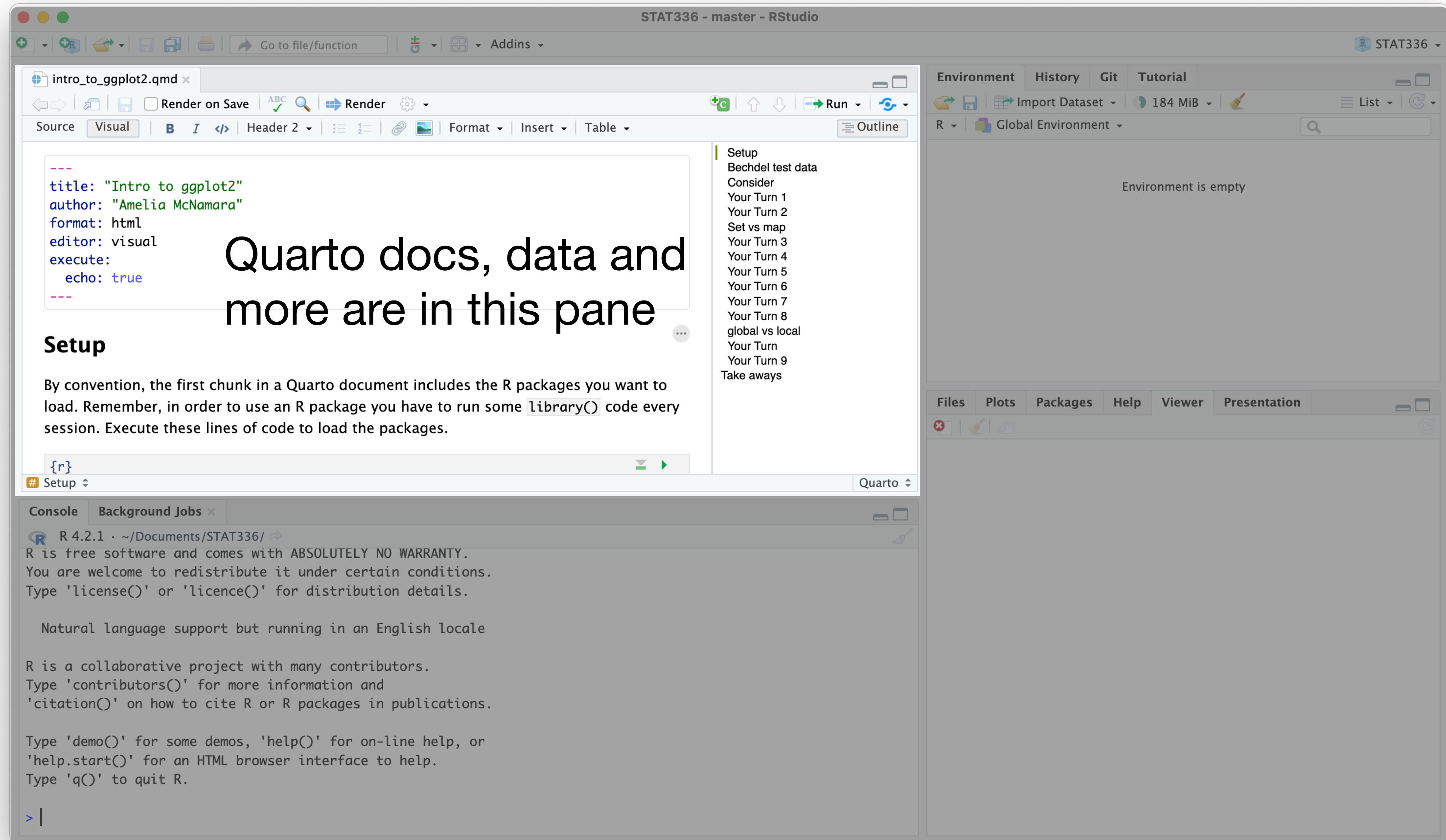
RStudio



RStudio



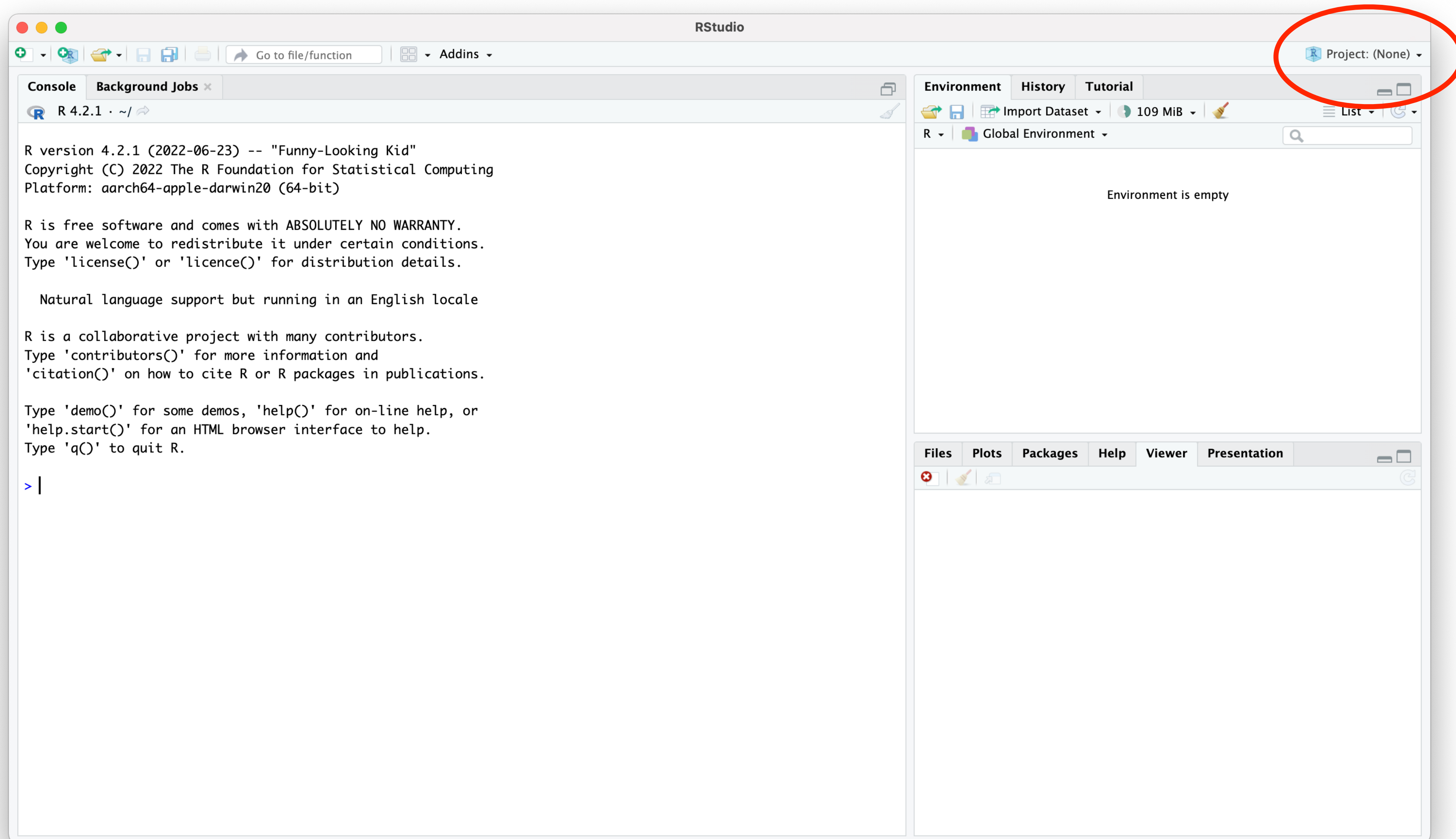
RStudio



RStudio projects



I highly, highly recommend projects

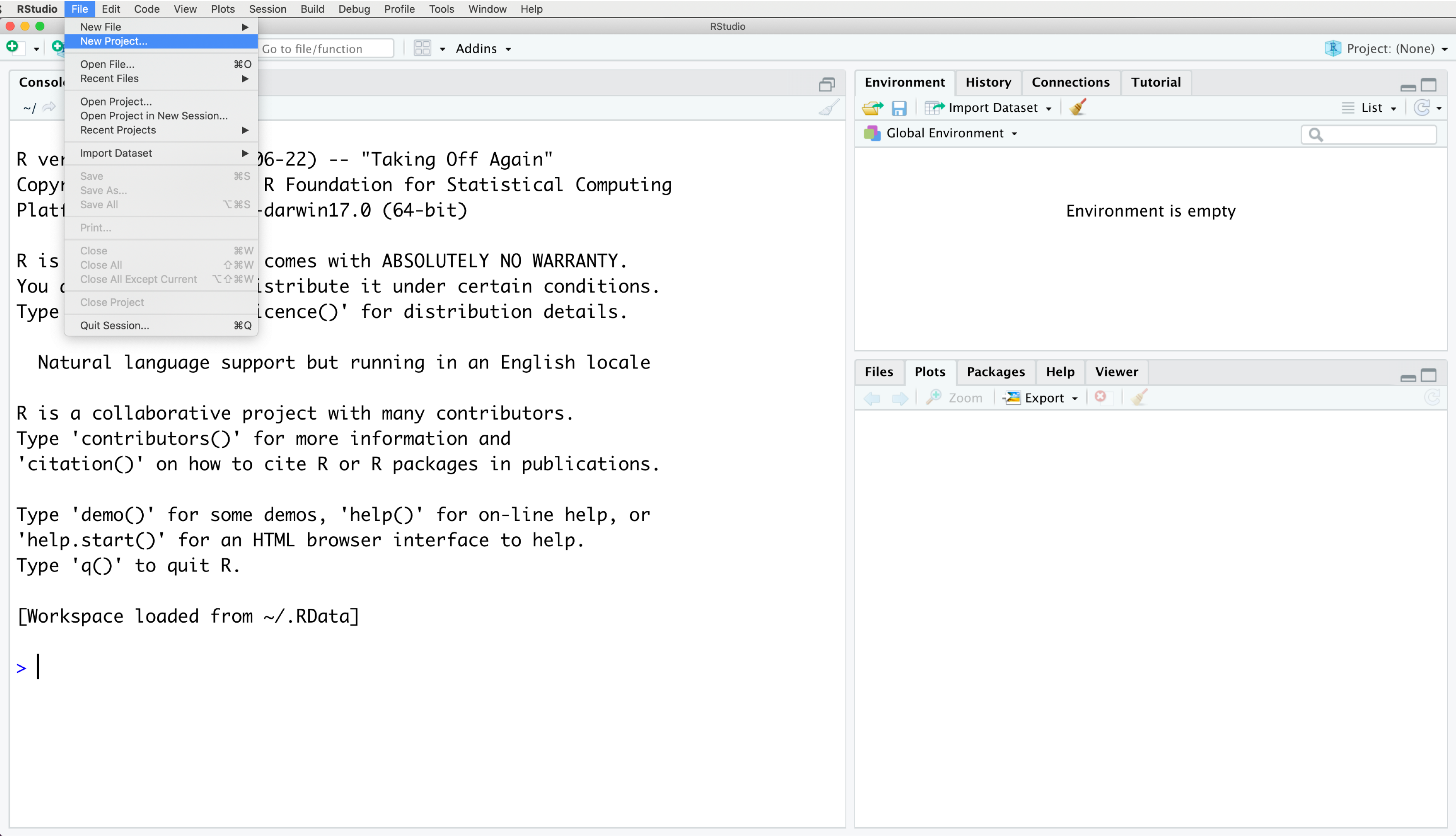


No project :(

Make a project

- File ->
- New Project ->
- New Directory* ->
- New Project ->
- Pick a directory name (maybe DASC336, all one word) ->
- Pick where the directory will live (I **highly** suggest OneDrive) ->
- Create Project
- RStudio should restart and you should then see the name of your project in the upper right corner of RStudio.

*Unless you already have a folder created for this class



Console

Terminal ×

Jobs ×



Environment

History

Connections

Tutorial



~/



Import Dataset ▾



List ▾



Global Environment ▾



R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOL
You are welcome to redistribute it unde
Type 'license()' or 'licence()' for dis

Natural language support but running

R is a collaborative project with many
Type 'contributors()' for more informat
'citation()' on how to cite R or R pack

Type 'demo()' for some demos, 'help()'
'help.start()' for an HTML browser inte
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> |

Environment is empty

New Project Wizard

Create Project



New Directory

Start a project in a brand new working directory



Existing Directory

Associate a project with an existing working directory



Version Control

Checkout a project from a version control repository



Cancel

Help

Viewer



Export ▾



STAT336 - master - RStudio

Go to file/function

Addins

intro_to_ggplot2.qmd

Render on Save

Render

Run

Source

Visual

B

I

</>

Header 2

Format

Insert

Table

Outline

title: "Intro to ggplot2"

author: "Amelia McNamara"

format: html

editor: visual

execute:

echo: true

Setup

Bechdel test data

Consider

Your Turn 1

Your Turn 2

Set vs map

Your Turn 3

Your Turn 4

Your Turn 5

Your Turn 6

Your Turn 7

Your Turn 8

global vs local

Your Turn

Your Turn 9

Take aways

{r}

Setup

Quarto

Console

Background Jobs

R 4.2.1 · ~/Documents/STAT336/

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Environment

History

Git

Tutorial

Import Dataset

184 MiB

List

R

Global Environment

Environment is empty

Files

Plots

Packages

Help

Viewer

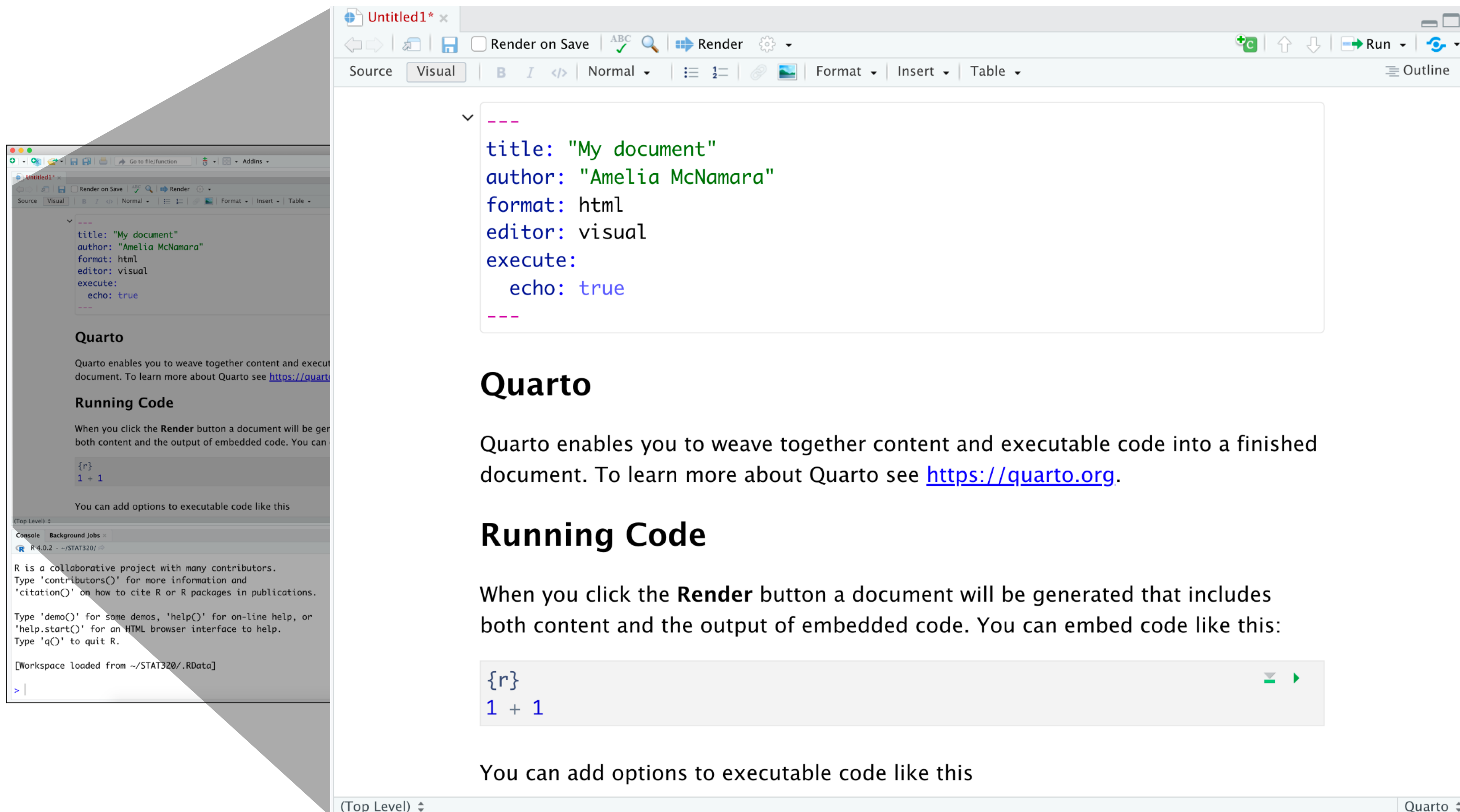
Presentation

STAT336

Yay!

Quarto

An authoring format for Data Science.



RStudio settings

- The default settings in RStudio are mostly pretty good! But, there are a couple I like to change. Let's do that. Go to:
- Tools ->
- Global Options
- On the General page, change "Save workspace to .RData on exit" to "Never" On the RMarkdown page, uncheck "Show output inline in RMarkdown documents" If you prefer a different color scheme, check out the Appearance page
- When you're done setting things up, hit Okay.

ggplot2



bechdel

Data on movies and the Bechdel test

```
library(tidyverse)
library(fivethirtyeight)
data(bechdel)
bechdel
?bechdel
```

Consider

Confer with the people around you.

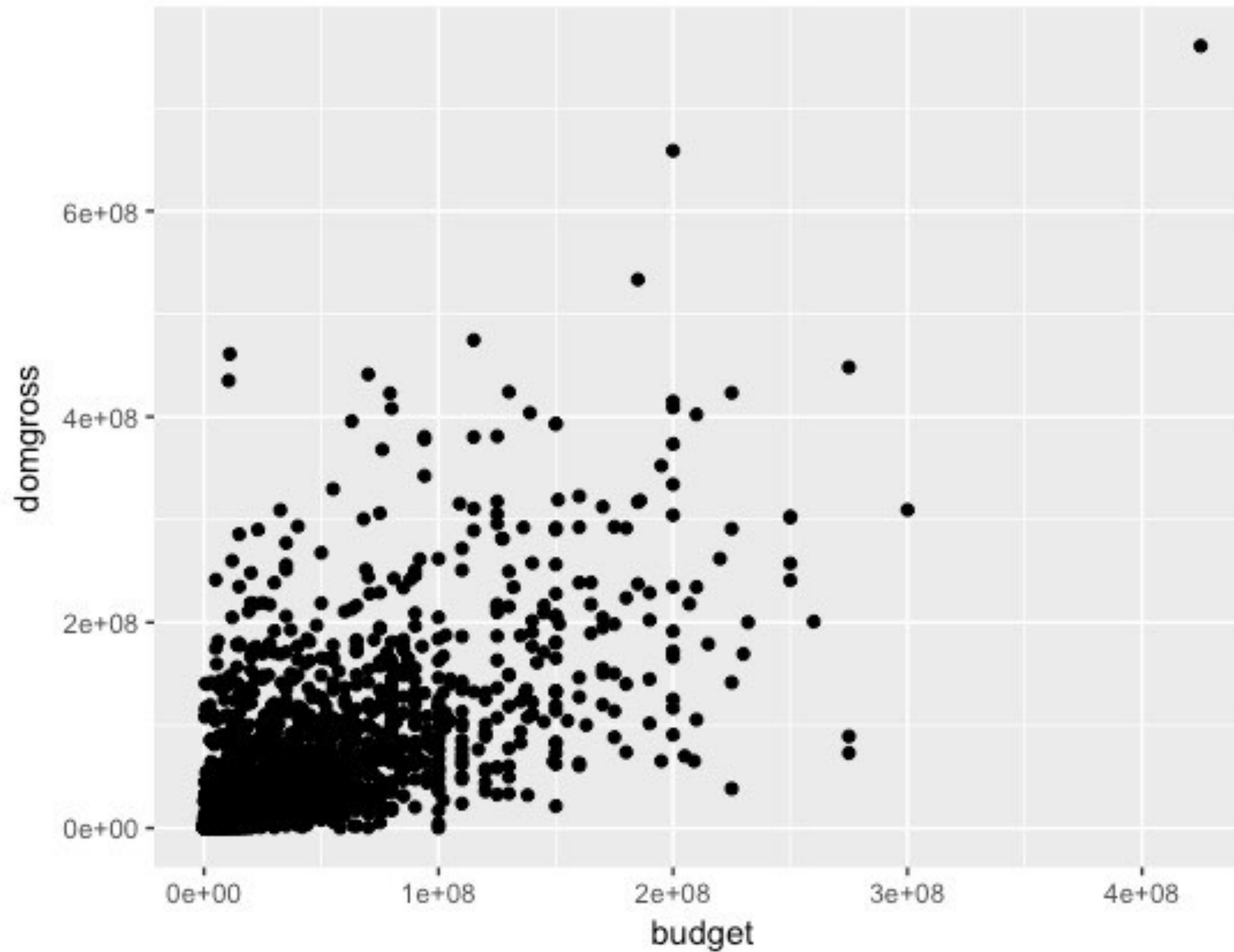
What relationship do you expect to see between movie budget (budget) and domestic gross(domgross)?

Your Turn 1

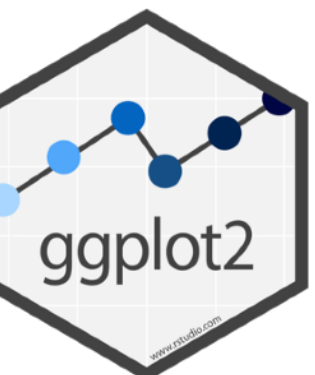
Run this code in your document to make a graph.

Pay strict attention to spelling, capitalization, and parentheses!

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```



```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

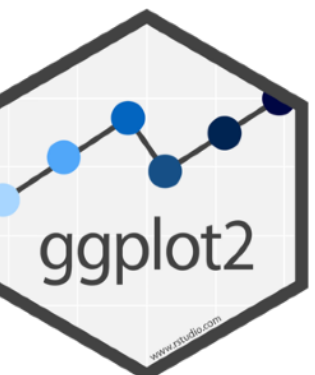


When you run this code, you will get what looks like an error, but is actually just a message from R. Some of the rows in the dataset didn't contain information for budget and/or domgross, so they're not plotted.



Removed 17 rows containing missing values (geom_point).

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```



1. "Initialize" a plot with `ggplot()`
2. Add layers with `geom_` functions

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

Pro tip: Always put the +
at the end of a line,
never at the start

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

data

+ before new line

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

type of layer

aes()

x variable

y variable

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

A template

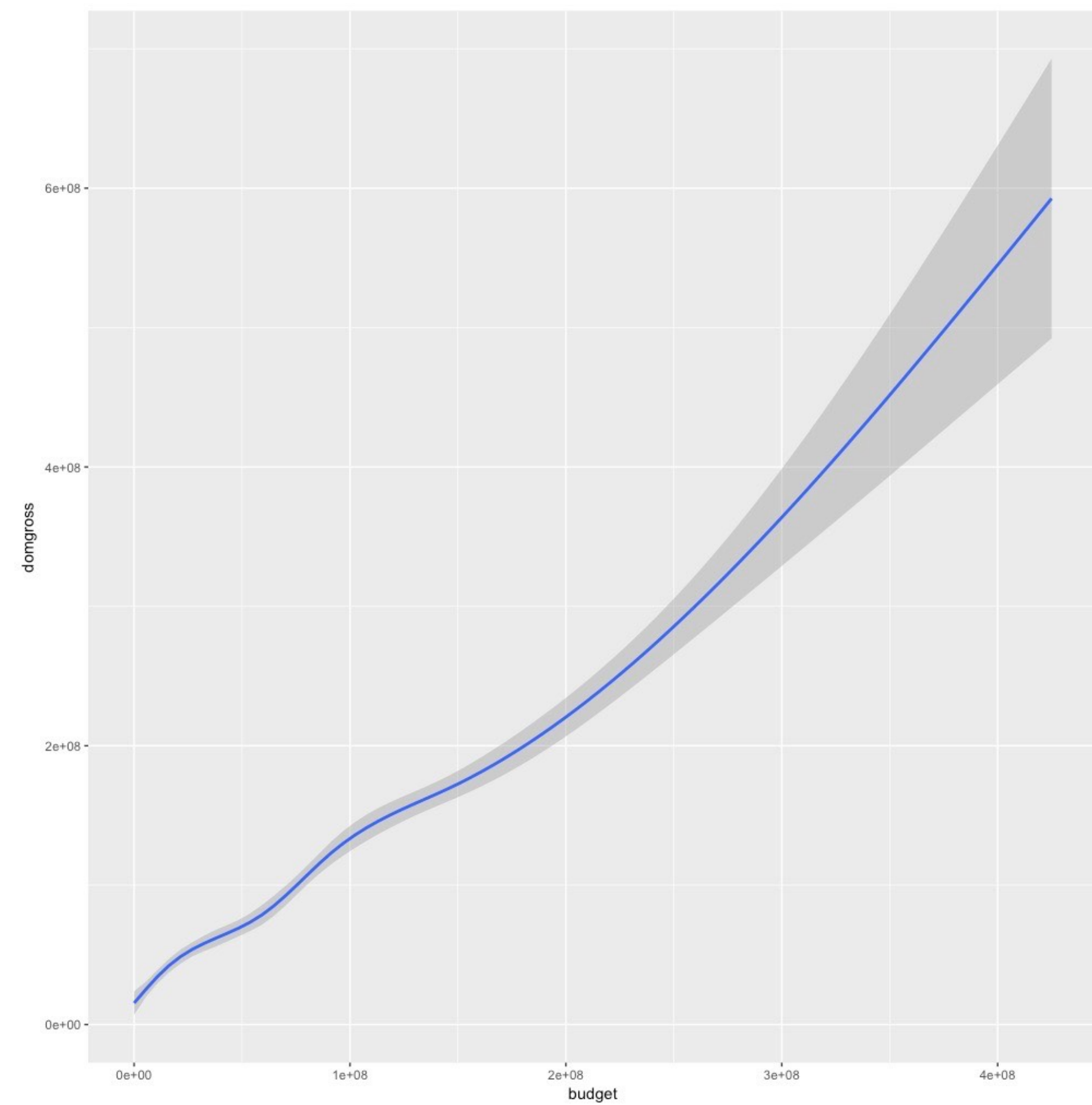
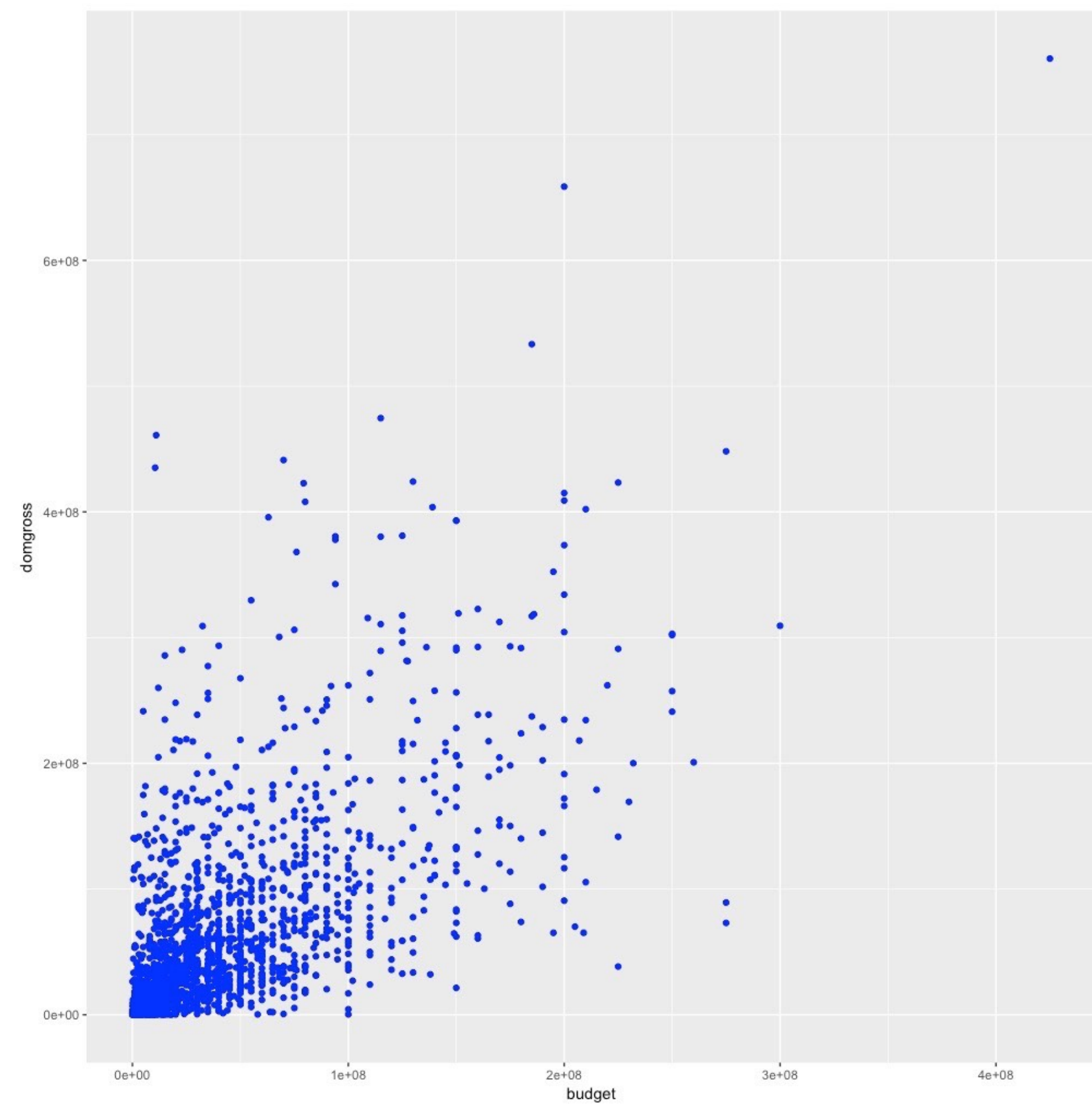
```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = budget, y = domgross))
```

geoms

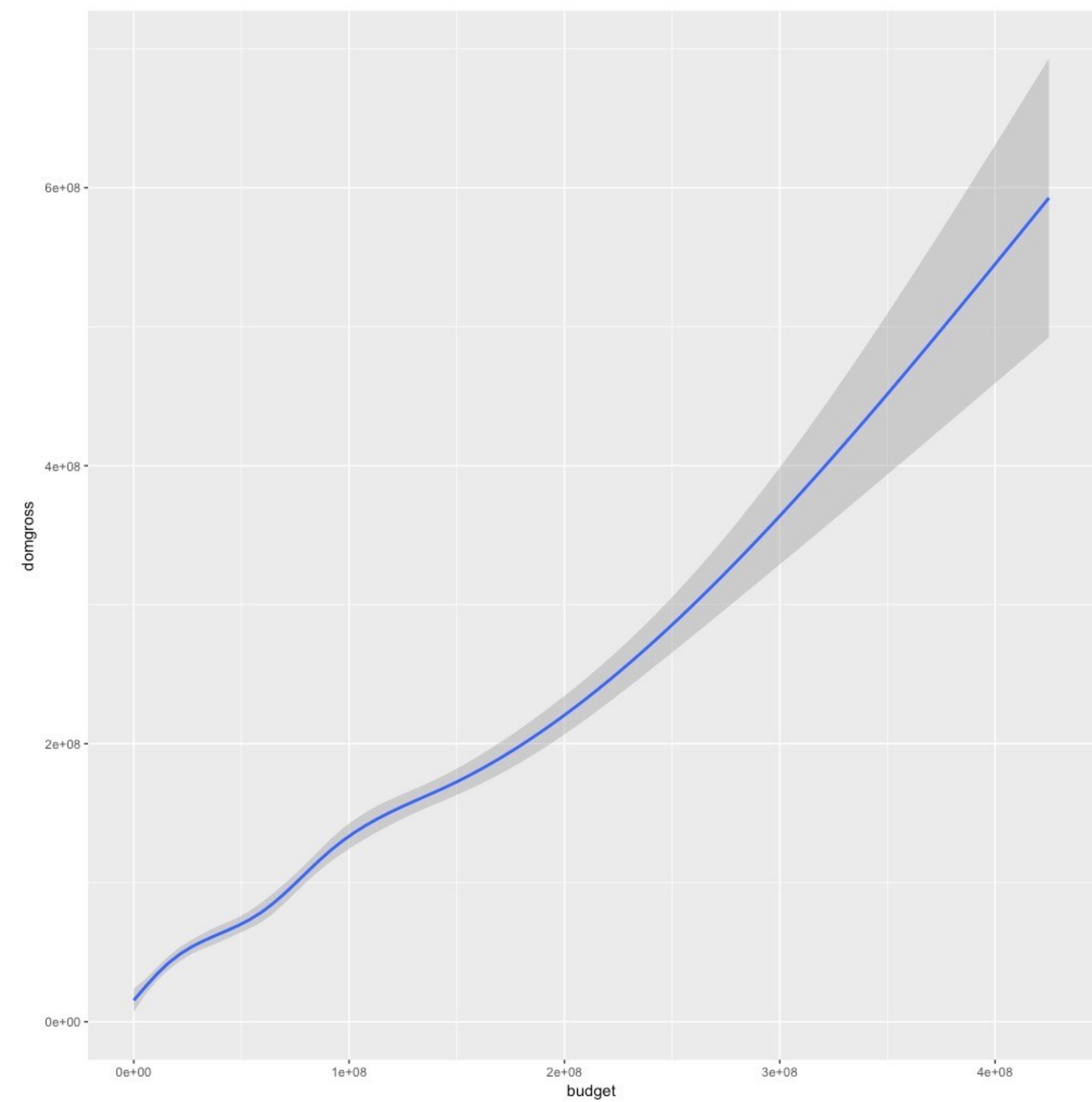
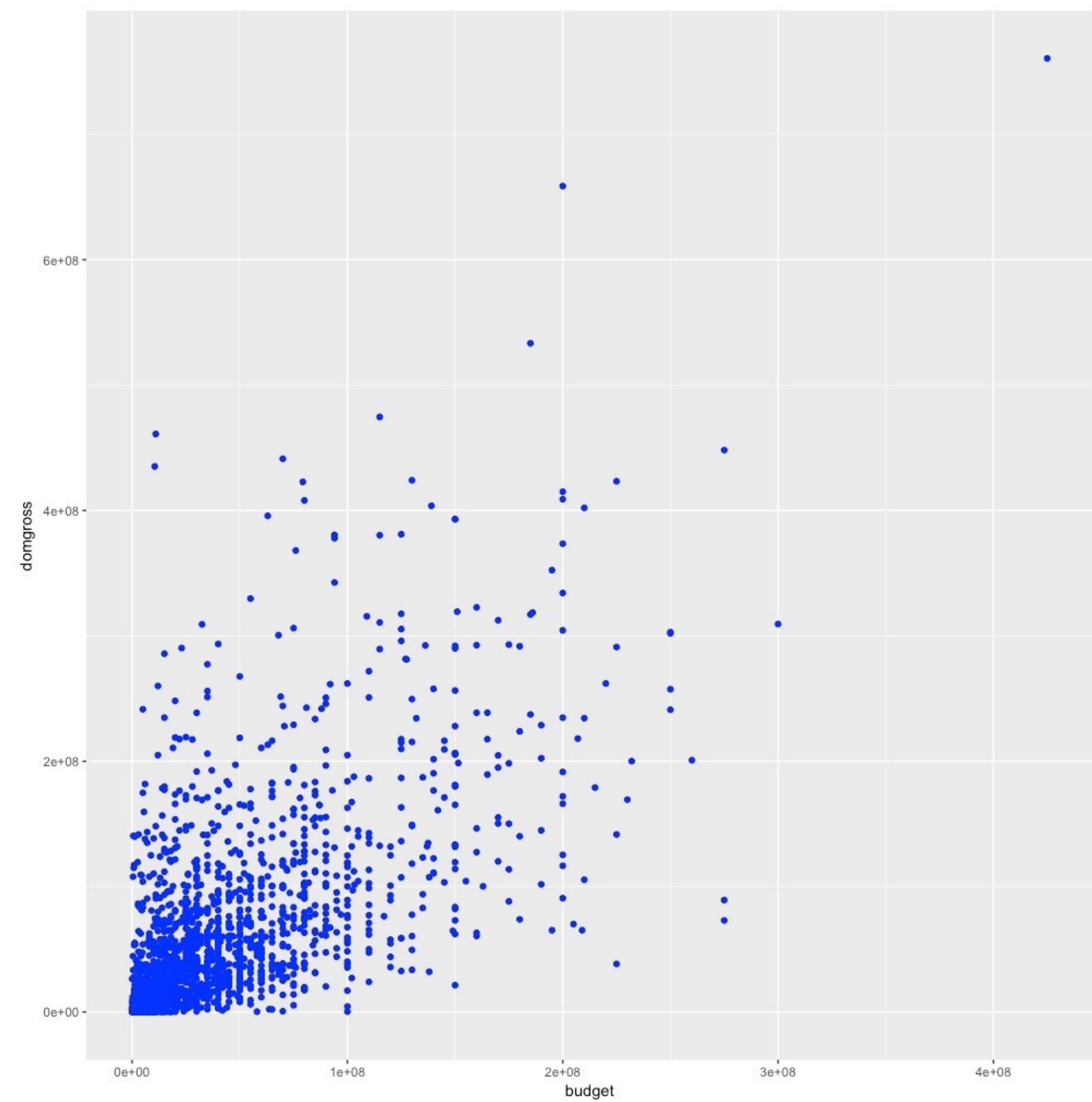


How are these plots similar?



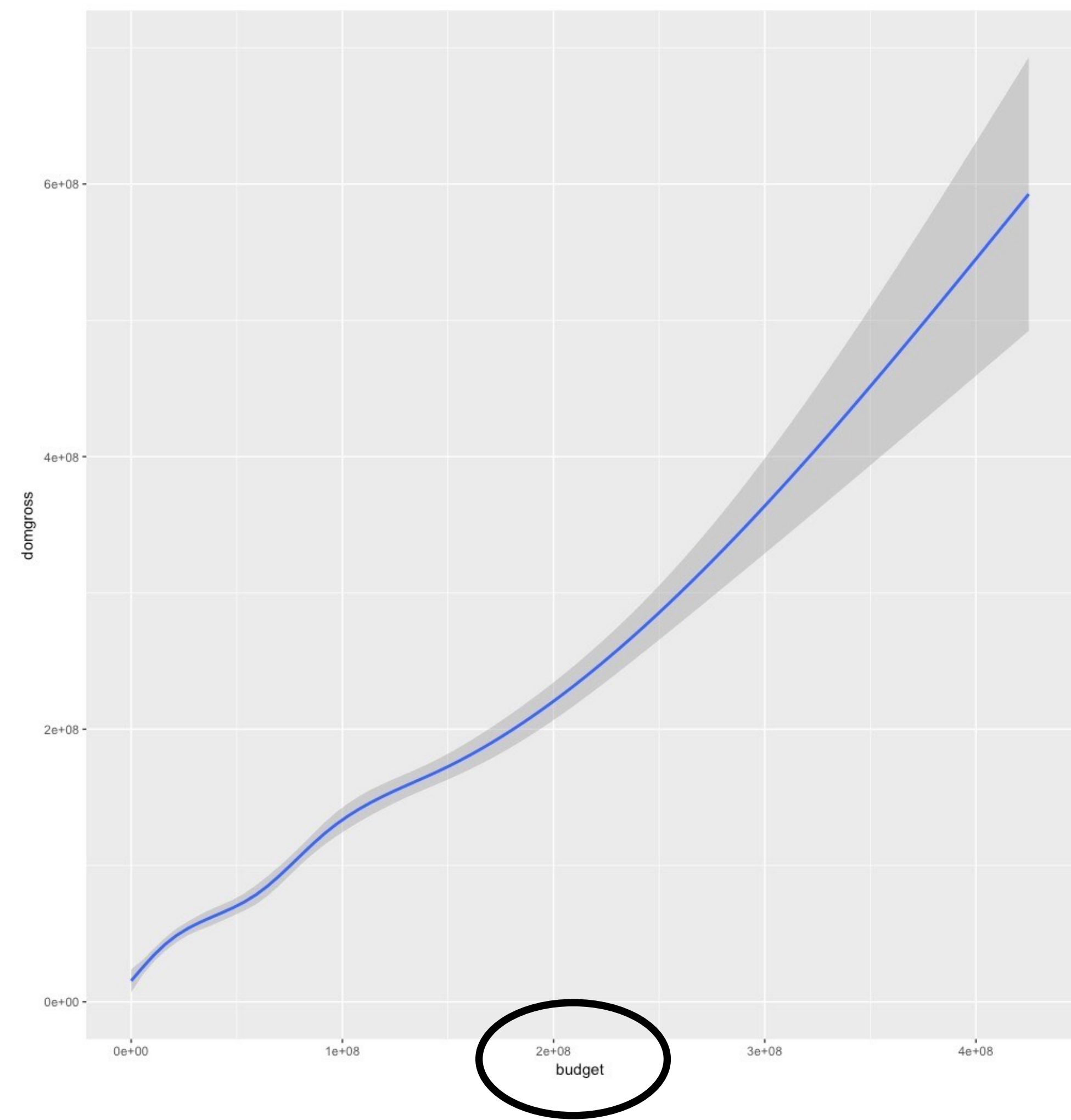
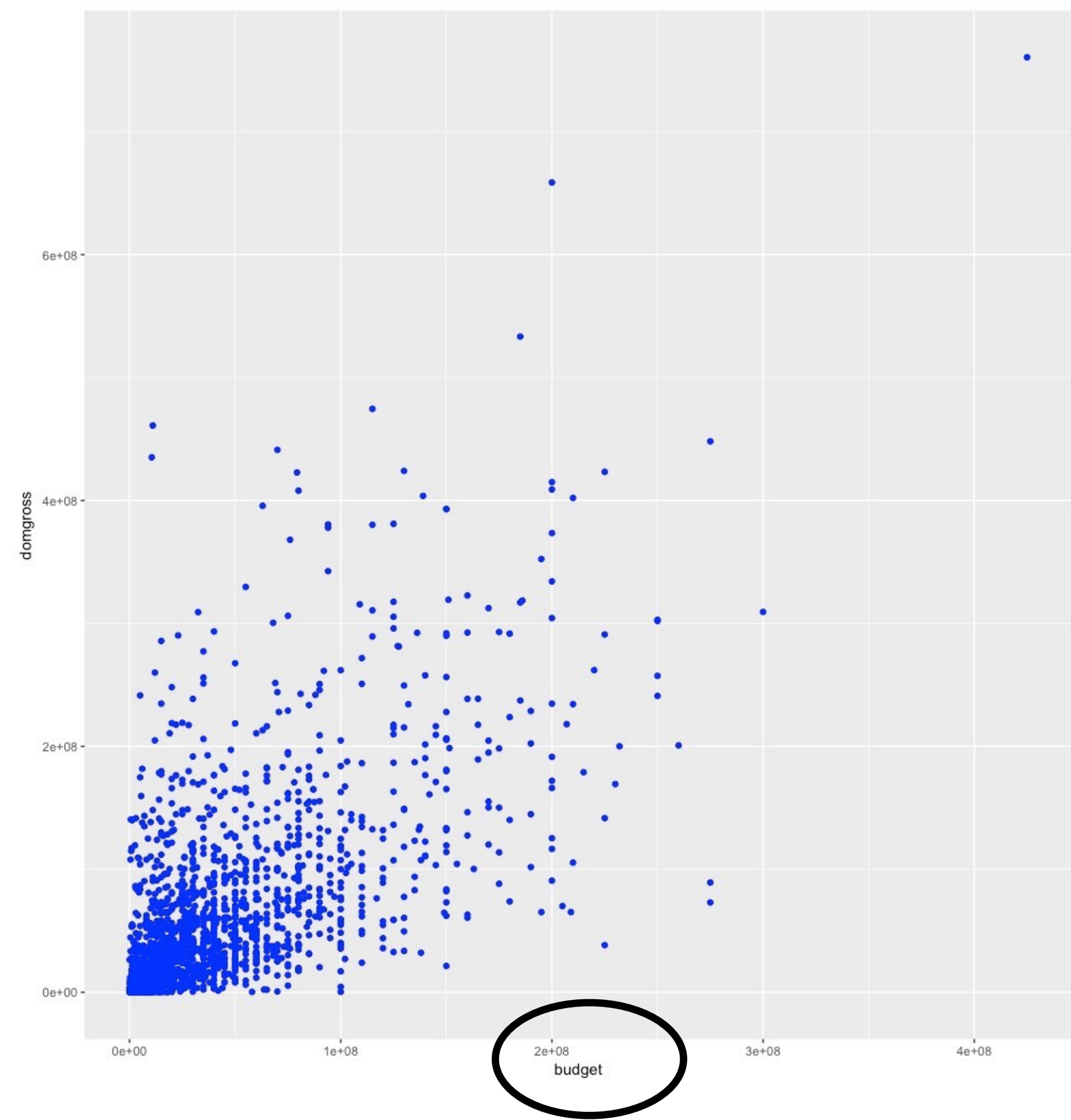
How are these plots similar?

Same:



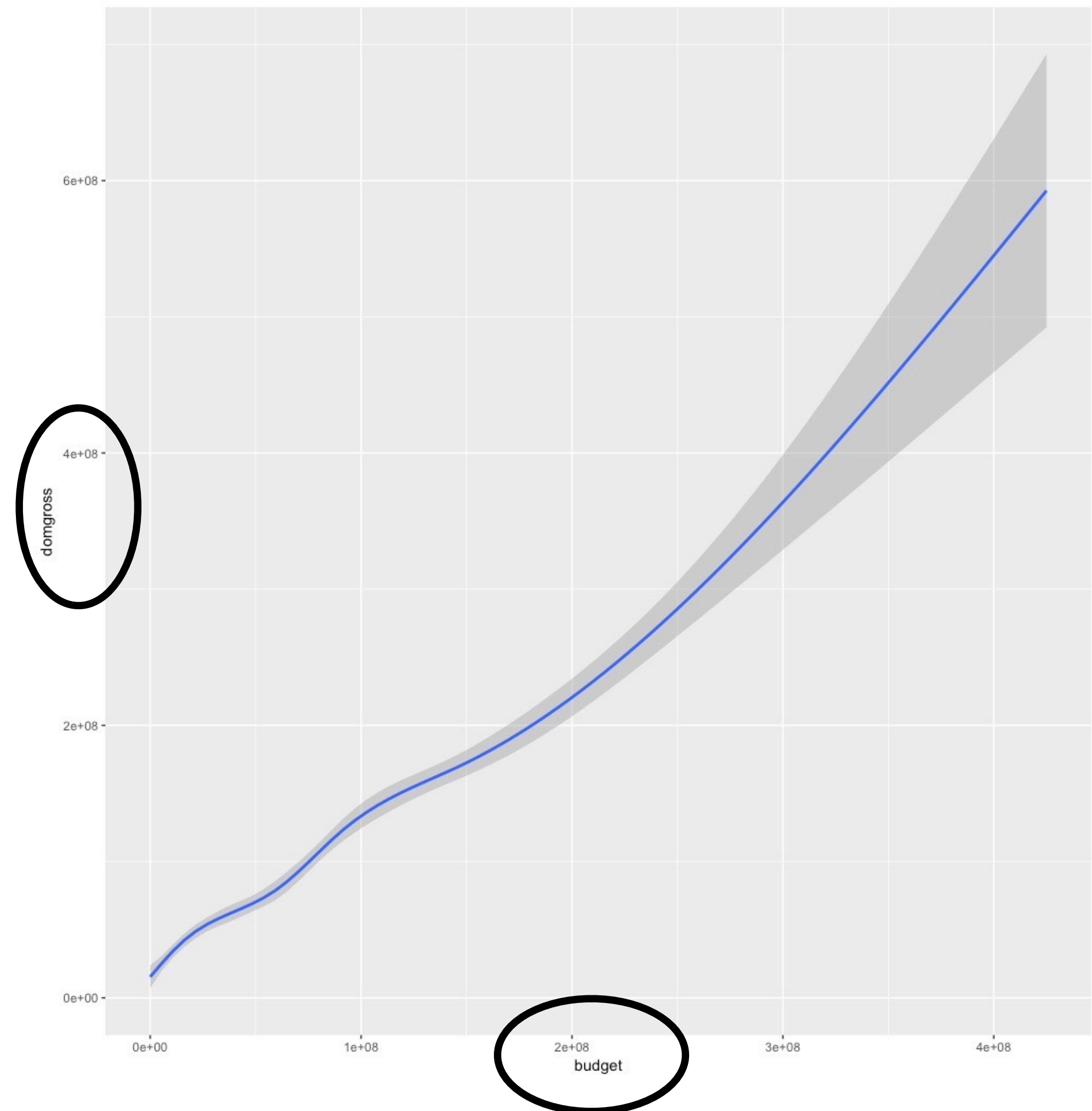
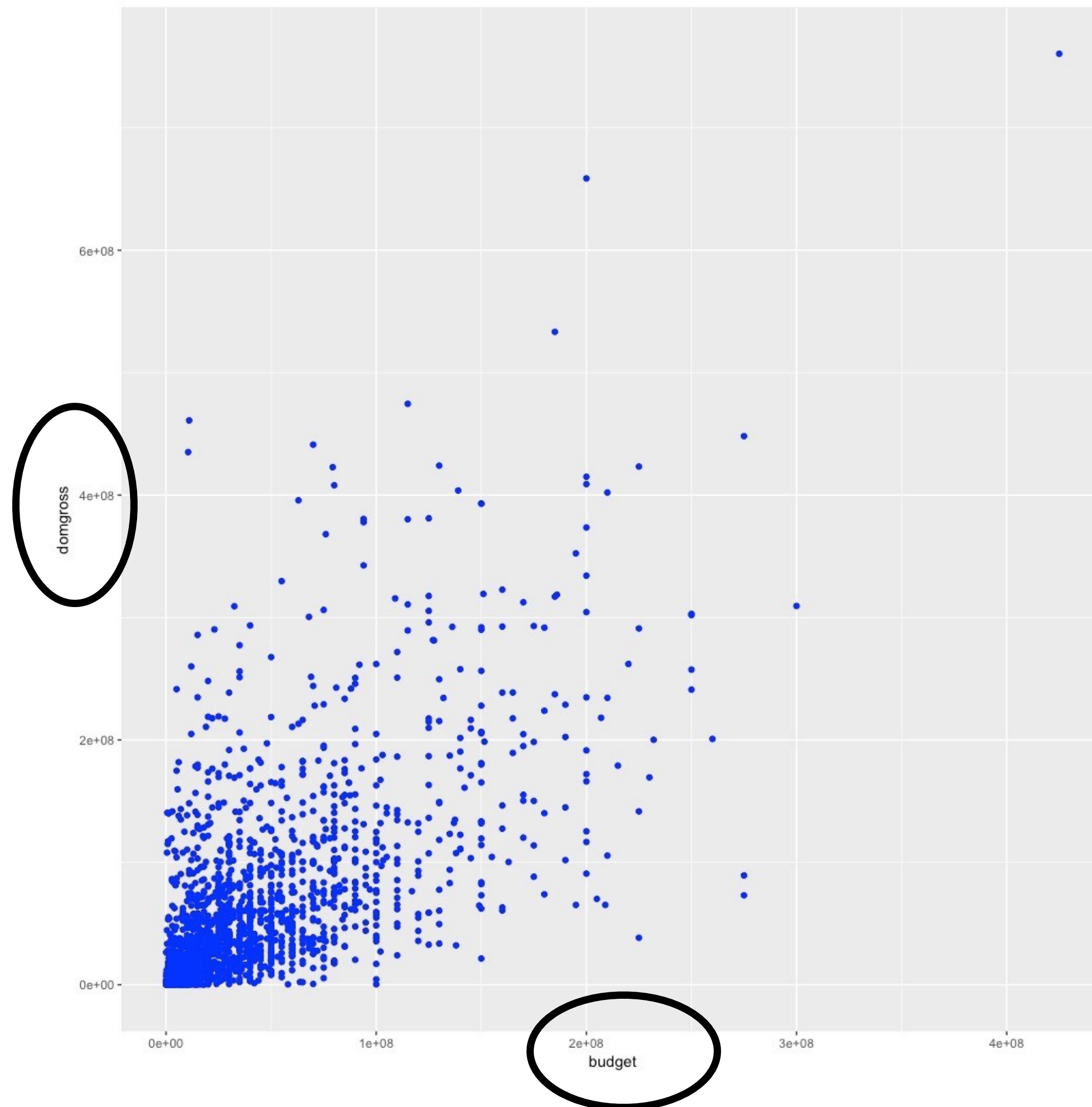
How are these plots similar?

Same: x var



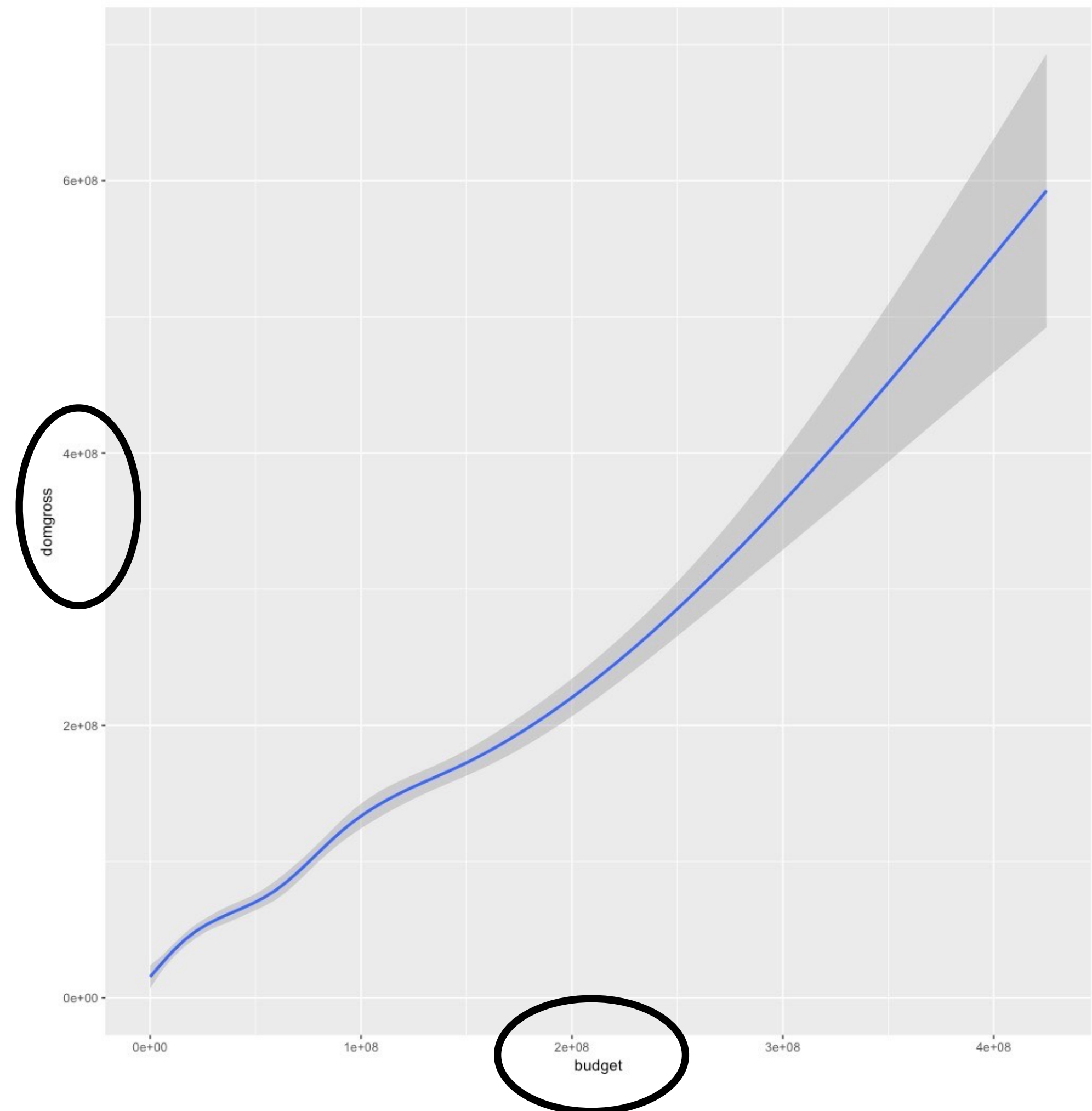
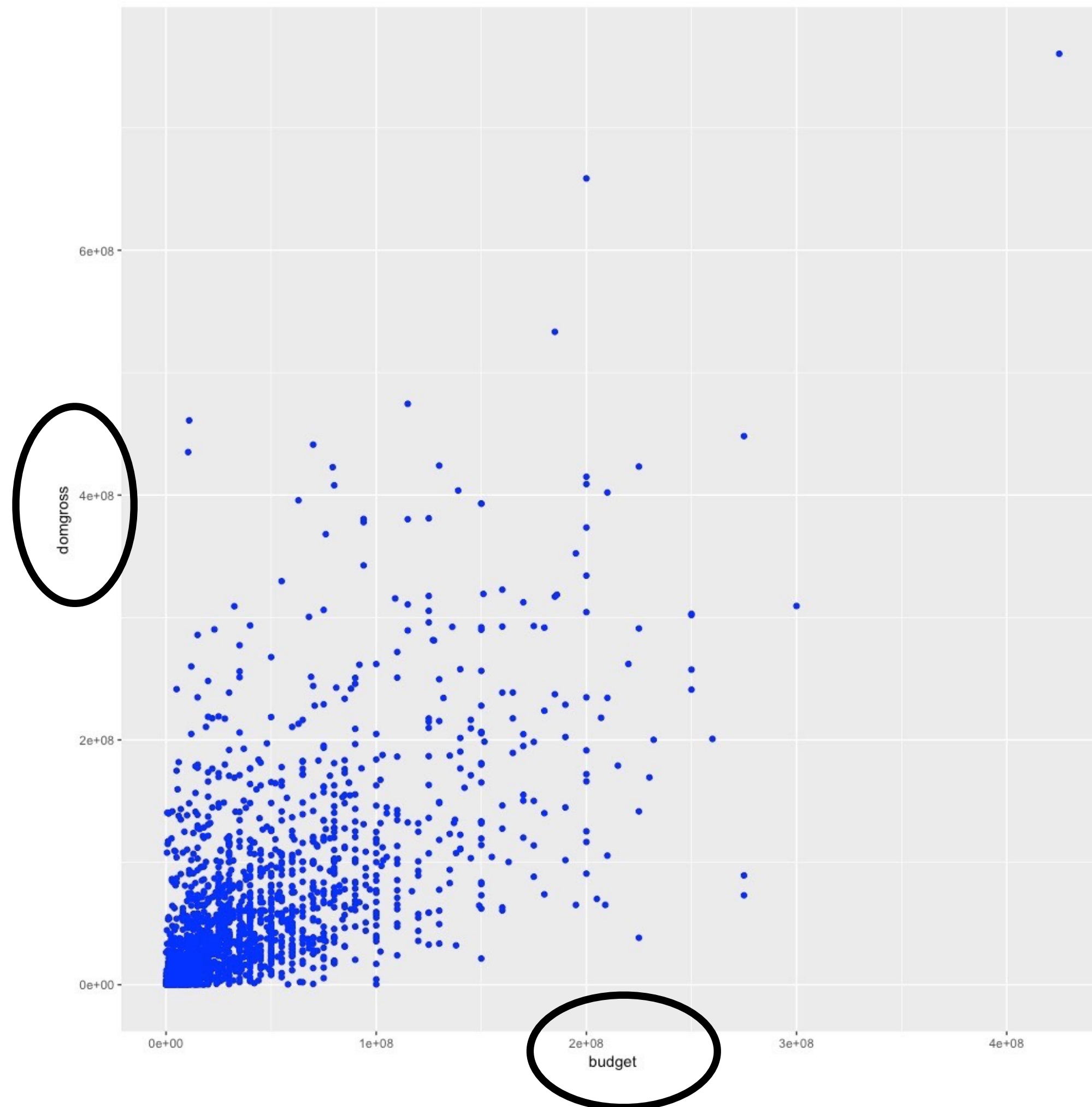
How are these plots similar?

Same: x var , y var

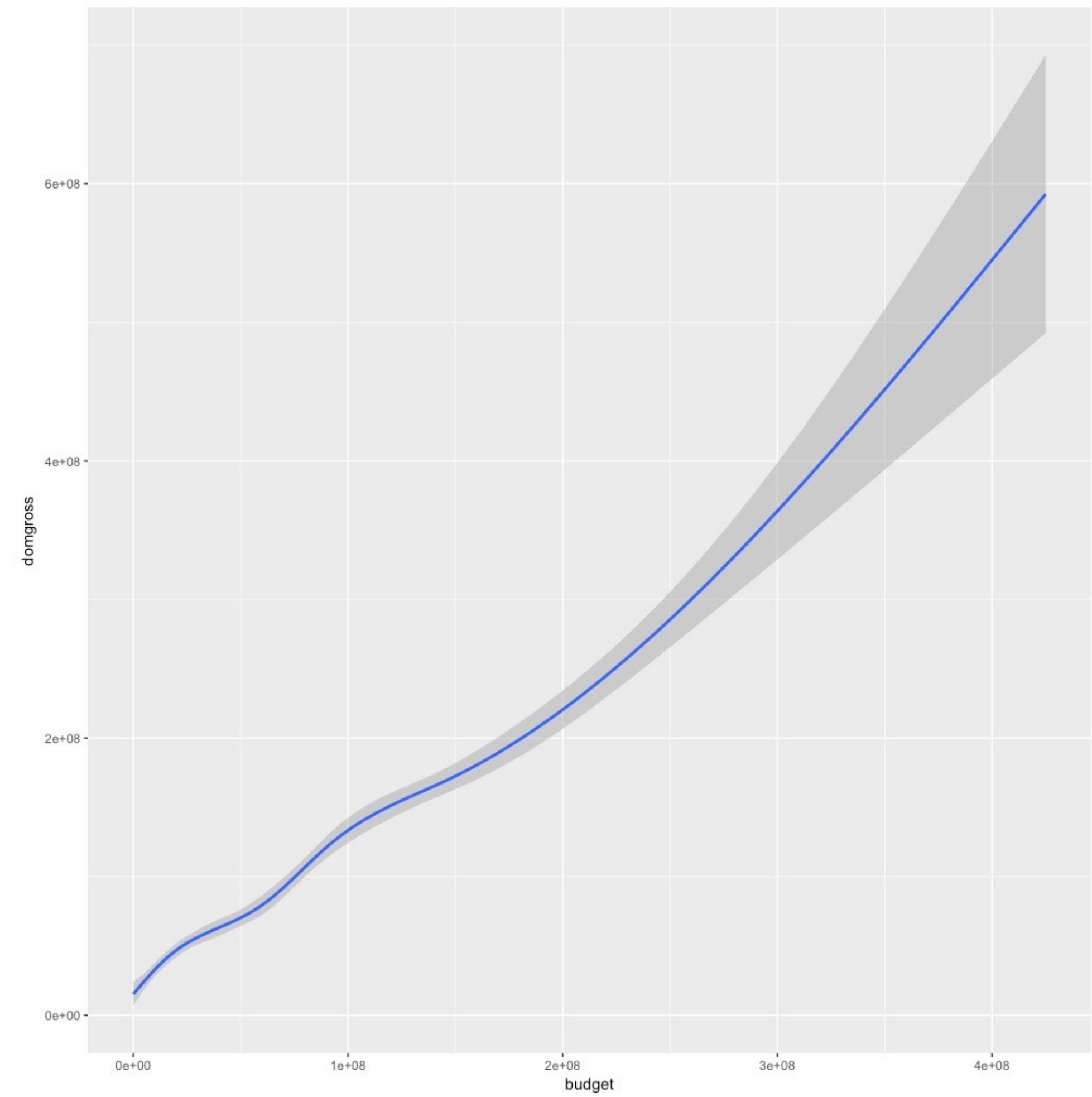
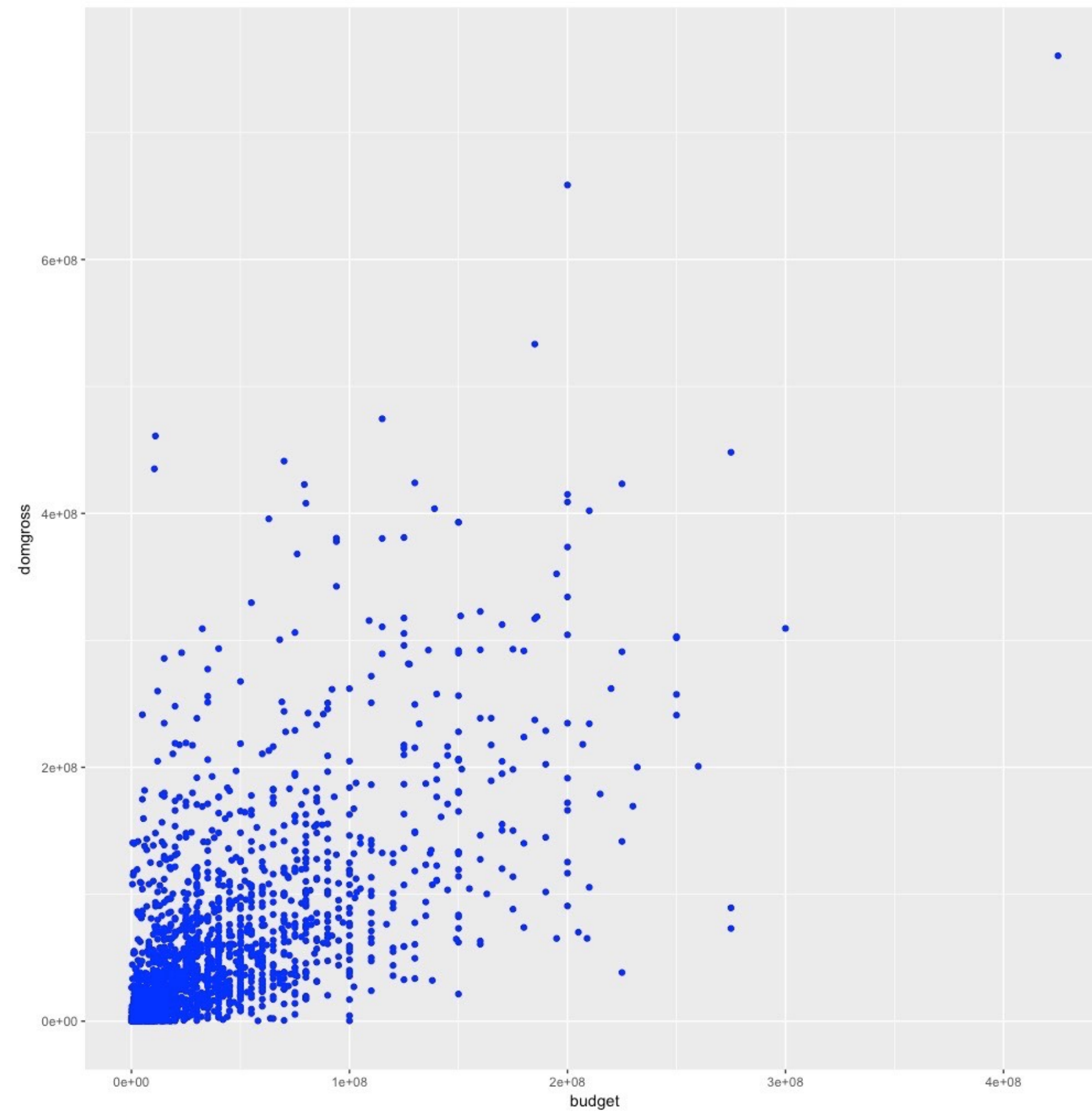


How are these plots similar?

Same: x var , y var , data

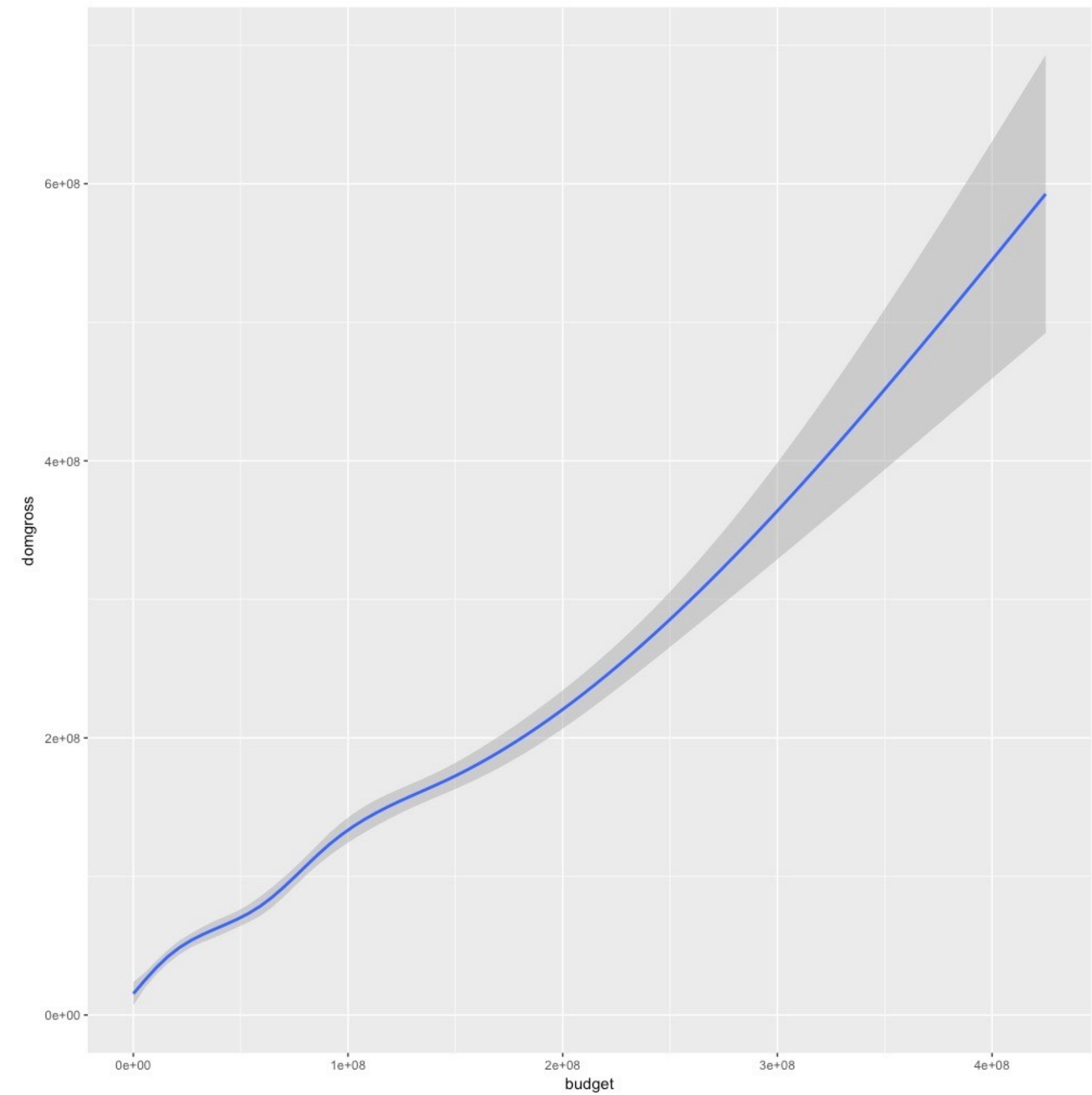
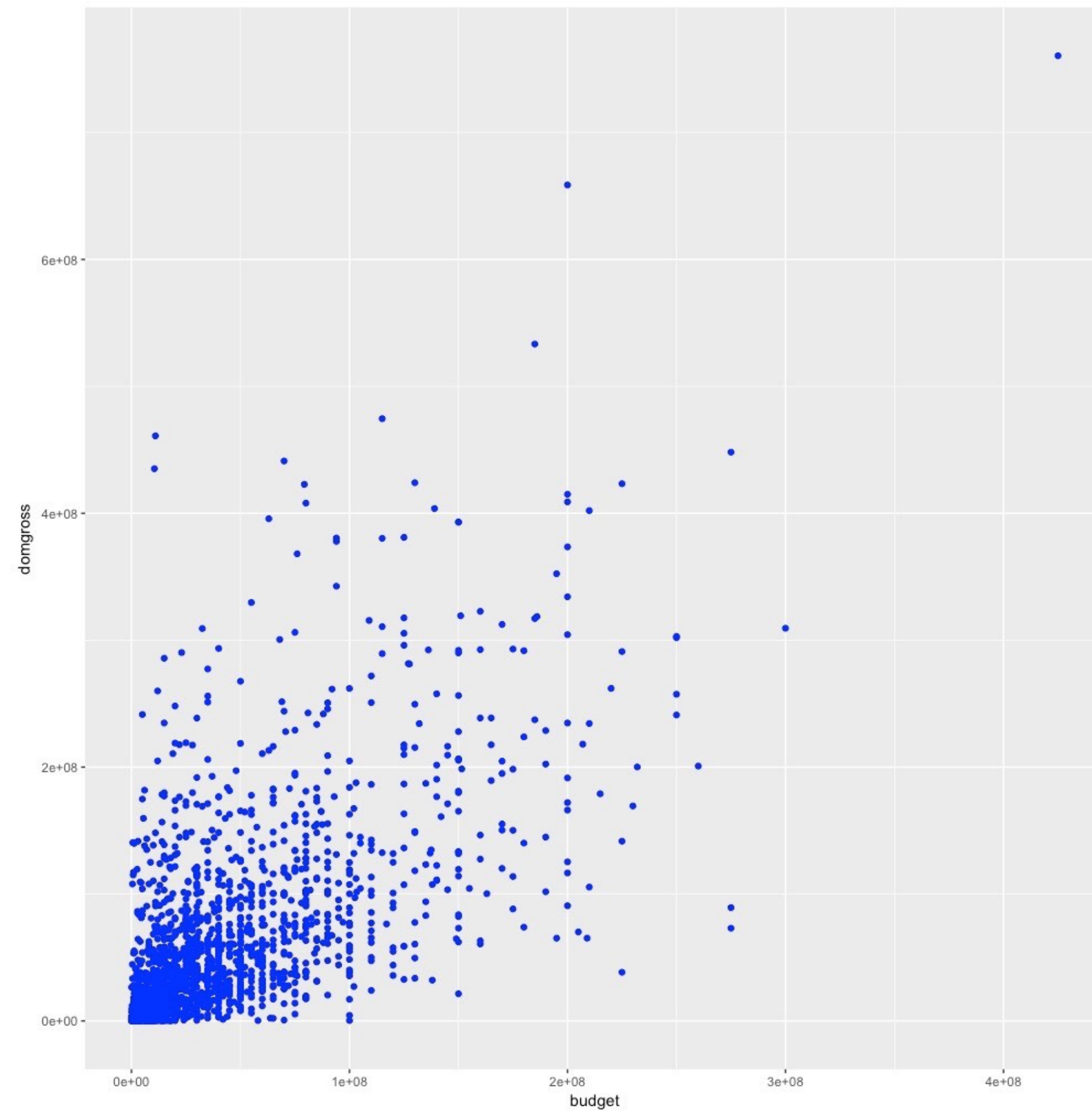


How are these plots different?



How are these plots different?

Different: geometric object (geom),
e.g. the visual object used to represent the data

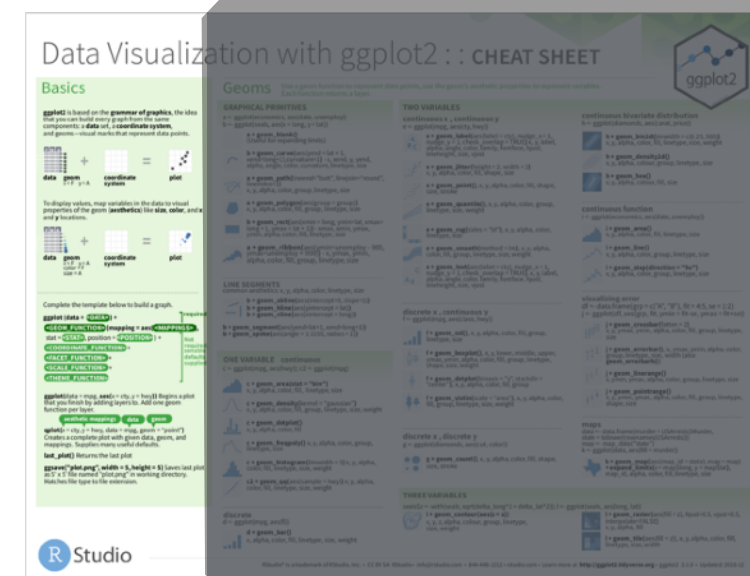


geoms

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

geom_ functions

Each requires a mapping argument.



Geoms

Use a geom function to represent data points, use the geom's aesthetic properties to represent variables.
Each function returns a layer.

GRAPHICAL PRIMITIVES

```
a <- ggplot(economics, aes(date, unemployment))
b <- ggplot(seals, aes(x = long, y = lat))
```

- a + geom_blank()**
(Useful for expanding limits)
- b + geom_curve**(aes(yend = lat + 1, xend=long+1, curvature=1) - x, xend, y, yend, alpha, angle, color, curvature, linetype, size)
- a + geom_path**(lineend="butt", linejoin="round", linemitre=1)
x, y, alpha, color, group, linetype, size
- a + geom_polygon**(aes(group = group))
x, y, alpha, color, fill, group, linetype, size
- b + geom_rect**(aes(xmin = long, ymin=lat, xmax=long + 1, ymax = lat + 1)) - xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size
- a + geom_ribbon**(aes(ymin=unemploy - 900, ymax=unemploy + 900)) - x, ymax, ymin, alpha, color, fill, group, linetype, size

LINE SEGMENTS

common aesthetics: x, y, alpha, color, linetype, size

- b + geom_abline**(aes(intercept=0, slope=1))
- b + geom_hline**(aes(yintercept = lat))
- b + geom_vline**(aes(xintercept = long))
- b + geom_segment**(aes(yend=lat+1, xend=long+1))
- b + geom_spoke**(aes(angle = 1:1155, radius = 1))

ONE VARIABLE continuous

```
c <- ggplot(mpg, aes(hwy)); c2 <- ggplot(mpg)
```

- c + geom_area**(stat = "bin")
x, y, alpha, color, fill, linetype, size
- c + geom_density**(kernel = "gaussian")
x, y, alpha, color, fill, group, linetype, size, weight
- c + geom_dotplot**()
x, y, alpha, color, fill
- c + geom_freqpoly**() x, y, alpha, color, group, linetype, size
- c + geom_histogram**(binwidth = 5) x, y, alpha, color, fill, linetype, size, weight
- c2 + geom_qq**(aes(sample = hwy)) x, y, alpha, color, fill, linetype, size, weight

discrete

```
d <- ggplot(mpg, aes(fl))
```

- d + geom_bar**()
x, alpha, color, fill, linetype, size, weight

TWO VARIABLES

continuous x , continuous y

```
e <- ggplot(mpg, aes(cty, hwy))
```

- e + geom_label**(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE) x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust
- e + geom_jitter**(height = 2, width = 2)
x, y, alpha, color, fill, shape, size
- e + geom_point**() x, y, alpha, color, fill, shape, size, stroke
- e + geom_quantile**() x, y, alpha, color, group, linetype, size, weight
- e + geom_rug**(sides = "bl") x, y, alpha, color, linetype, size
- e + geom_smooth**(method = lm) x, y, alpha, color, fill, group, linetype, size, weight
- e + geom_text**(aes(label = cty), nudge_x = 1, nudge_y = 1, check_overlap = TRUE) x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

discrete x , continuous y

```
f <- ggplot(mpg, aes(class, hwy))
```

- f + geom_col**() x, y, alpha, color, fill, group, linetype, size
- f + geom_boxplot**() x, y, lower, middle, upper, ymax, ymin, alpha, color, fill, group, linetype, shape, size, weight
- f + geom_dotplot**(binaxis = "y", stackdir = "center") x, y, alpha, color, fill, group
- f + geom_violin**(scale = "area") x, y, alpha, color, fill, group, linetype, size, weight

discrete x , discrete y

```
g <- ggplot(diamonds, aes(cut, color))
```

- g + geom_count**() x, y, alpha, color, fill, shape, size, stroke

THREE VARIABLES

```
sealsSz <- with(seals, sqrt(delta_long^2 + delta_lat^2)); l <- ggplot(seals, aes(long, lat))
```

- l + geom_contour**(aes(z = z))
x, y, z, alpha, colour, group, linetype, size, weight
- l + geom_raster**(aes(fill = z), hjust=0.5, vjust=0.5, interpolate=FALSE)
x, y, alpha, fill
- l + geom_tile**(aes(fill = z)) x, y, alpha, color, fill, linetype, size, width

continuous bivariate distribution

```
h <- ggplot(diamonds, aes(carat, price))
```

- h + geom_bin2d**(binwidth = c(0.25, 500))
x, y, alpha, color, fill, linetype, size, weight
- h + geom_density2d**()
x, y, alpha, colour, group, linetype, size
- h + geom_hex**()
x, y, alpha, colour, fill, size

continuous function

```
i <- ggplot(economics, aes(date, unemployment))
```

- i + geom_area**()
x, y, alpha, color, fill, linetype, size
- i + geom_line**()
x, y, alpha, color, group, linetype, size
- i + geom_step**(direction = "hv")
x, y, alpha, color, group, linetype, size

visualizing error

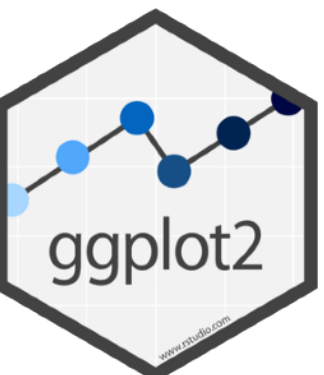
```
df <- data.frame(grp = c("A", "B"), fit = 4:5, se = 1:2)
j <- ggplot(df, aes(grp, fit, ymin = fit-se, ymax = fit+se))
```

- j + geom_crossbar**(fatten = 2)
x, y, ymax, ymin, alpha, color, fill, group, linetype, size
- j + geom_errorbar**() x, ymax, ymin, alpha, color, group, linetype, size, width (also **geom_errorbarh**())
- j + geom_linerange**()
x, ymin, ymax, alpha, color, group, linetype, size
- j + geom_pointrange**()
x, y, ymin, ymax, alpha, color, fill, group, linetype, shape, size

maps

```
data <- data.frame(murder = USArrests$Murder, state = tolower(rownames(USArrests)))
map <- map_data("state")
k <- ggplot(data, aes(fill = murder))
```

- k + geom_map**(aes(map_id = state), map = map) + **expand_limits**(x = map\$long, y = map\$lat, map_id, alpha, color, fill, linetype, size)

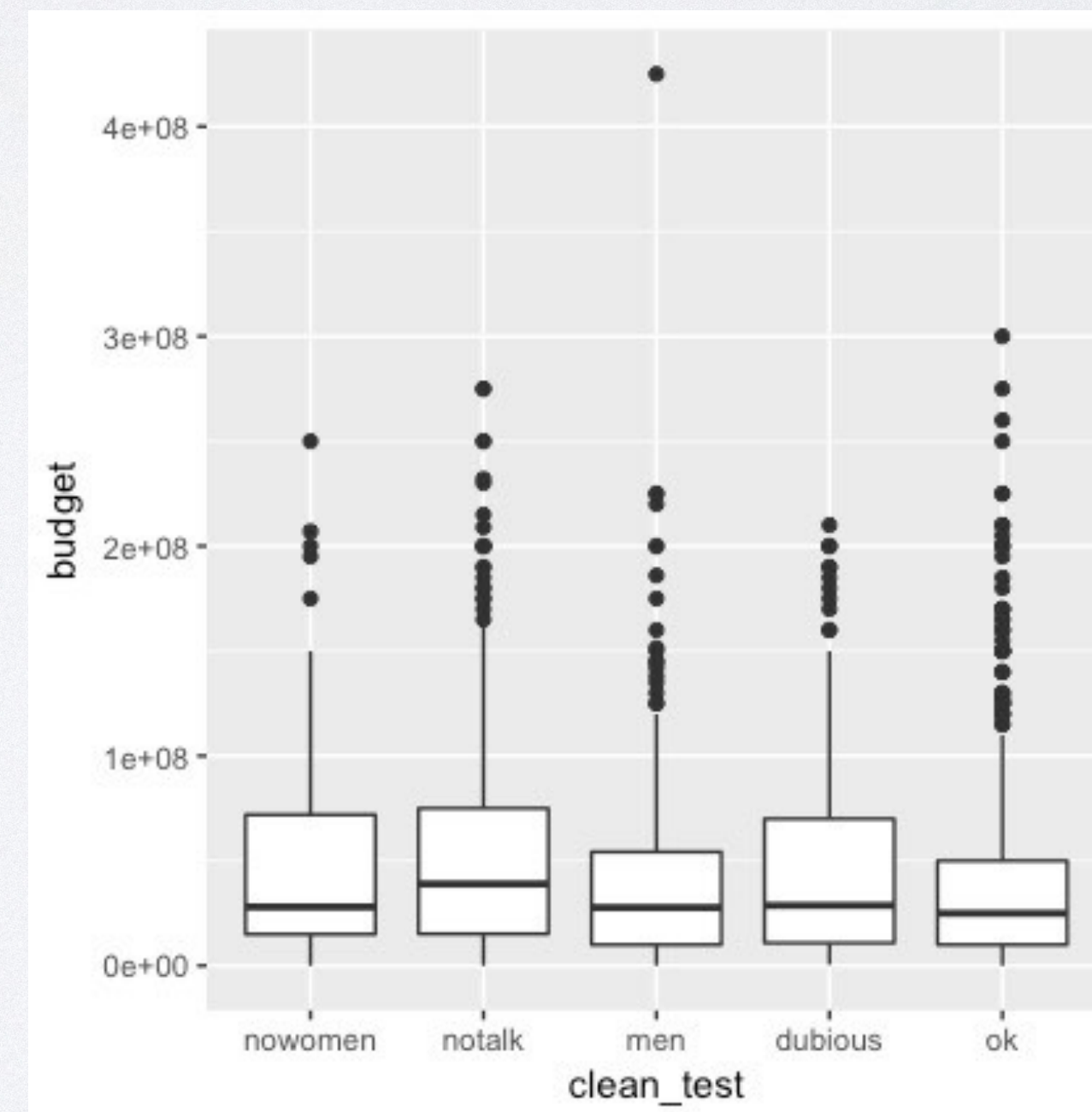
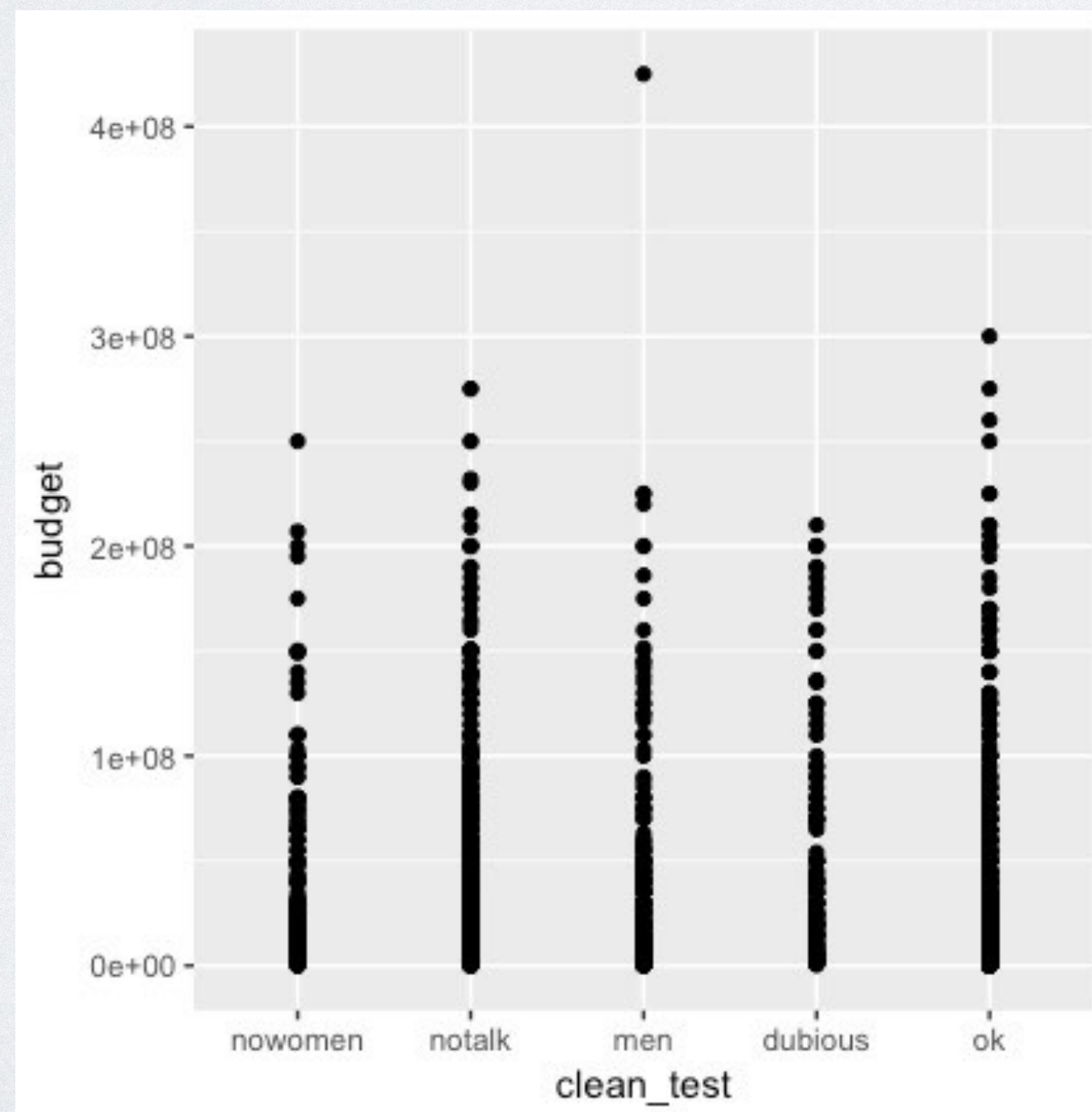


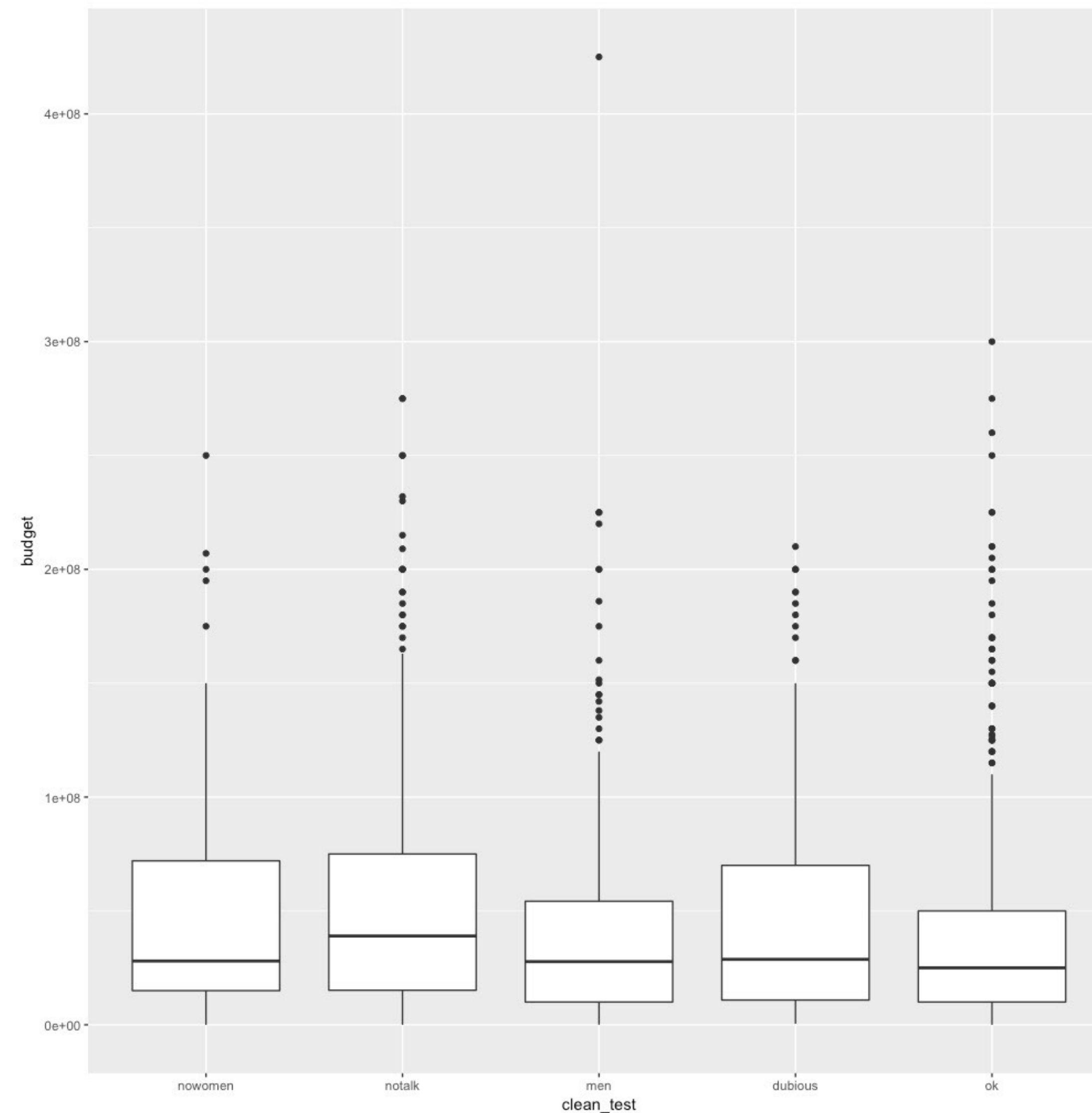
Your Turn 2

recall

```
ggplot(data = bechdel) +  
  geom_point(mapping = aes(x = clean_test, y = budget))
```

With your partner, decide how to replace this scatterplot with one that draws boxplots. Use the cheatsheet. Try your best guess.

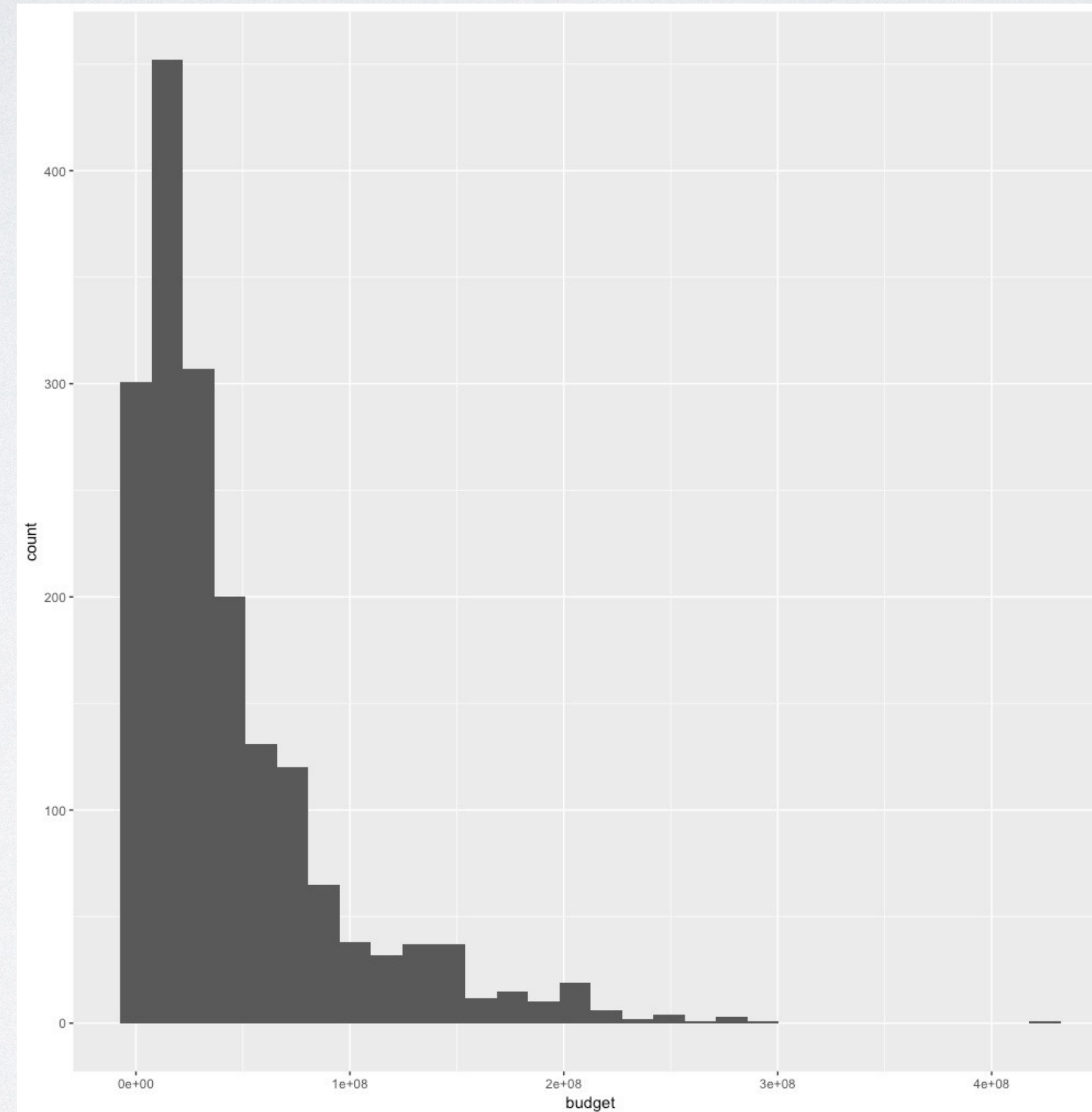


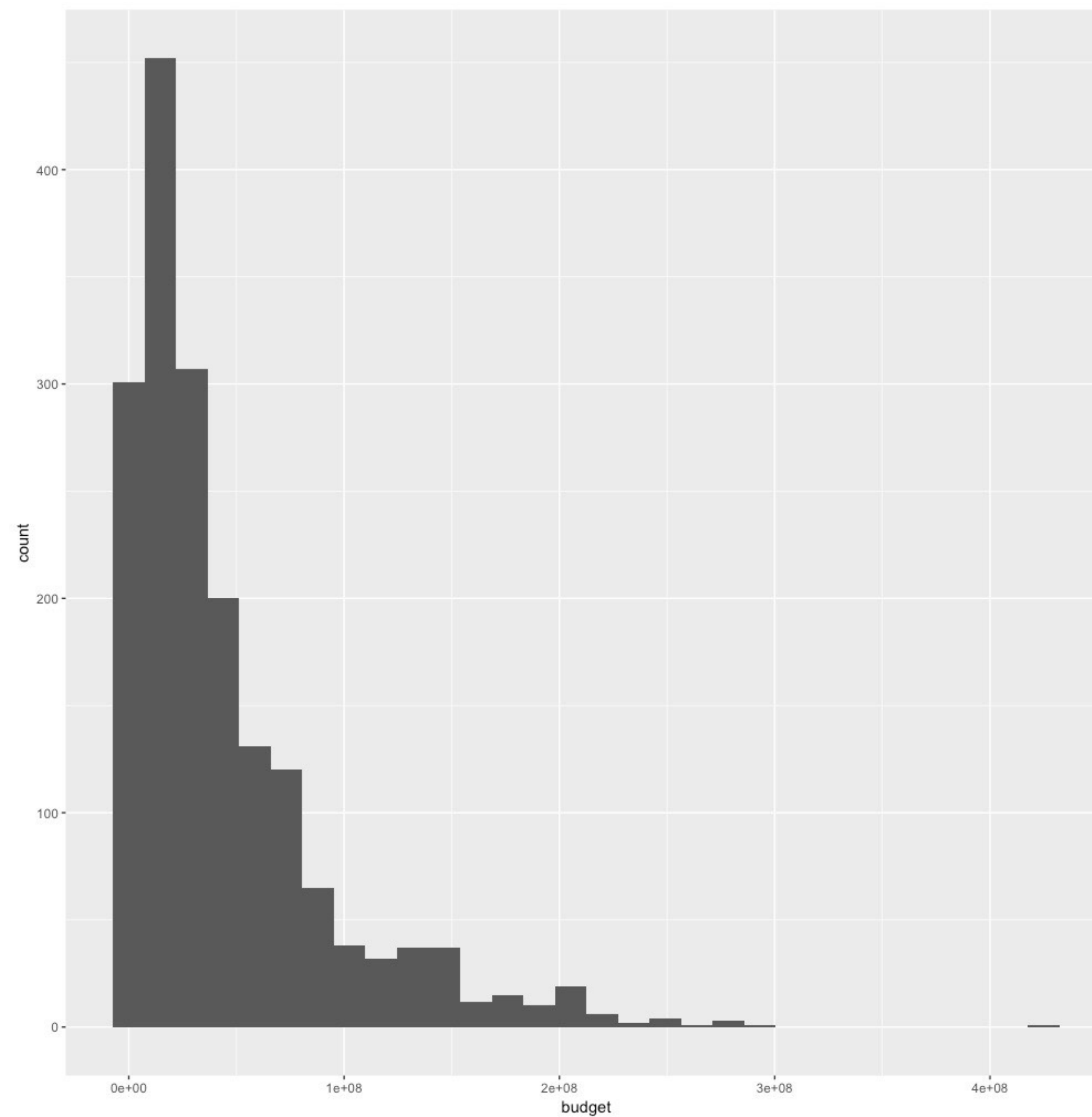


```
ggplot(data = bechdel) +  
  geom_boxplot(mapping = aes(x = clean_test, y = budget))
```

Your Turn 3

With your partner, make the histogram of **budget** below.
Use the cheatsheet. Hint: do not supply a **y** variable.





```
ggplot(data = bechdel) +  
  geom_histogram(mapping = aes(x = budget))
```

When you run this code, you will get what looks like an error, but is actually just a message from R. Because it's a bad idea to use default binwidths, the package is reminding you to pick your own



``stat_bin()` using `bins = 30`. Pick better value with `binwidth`.`

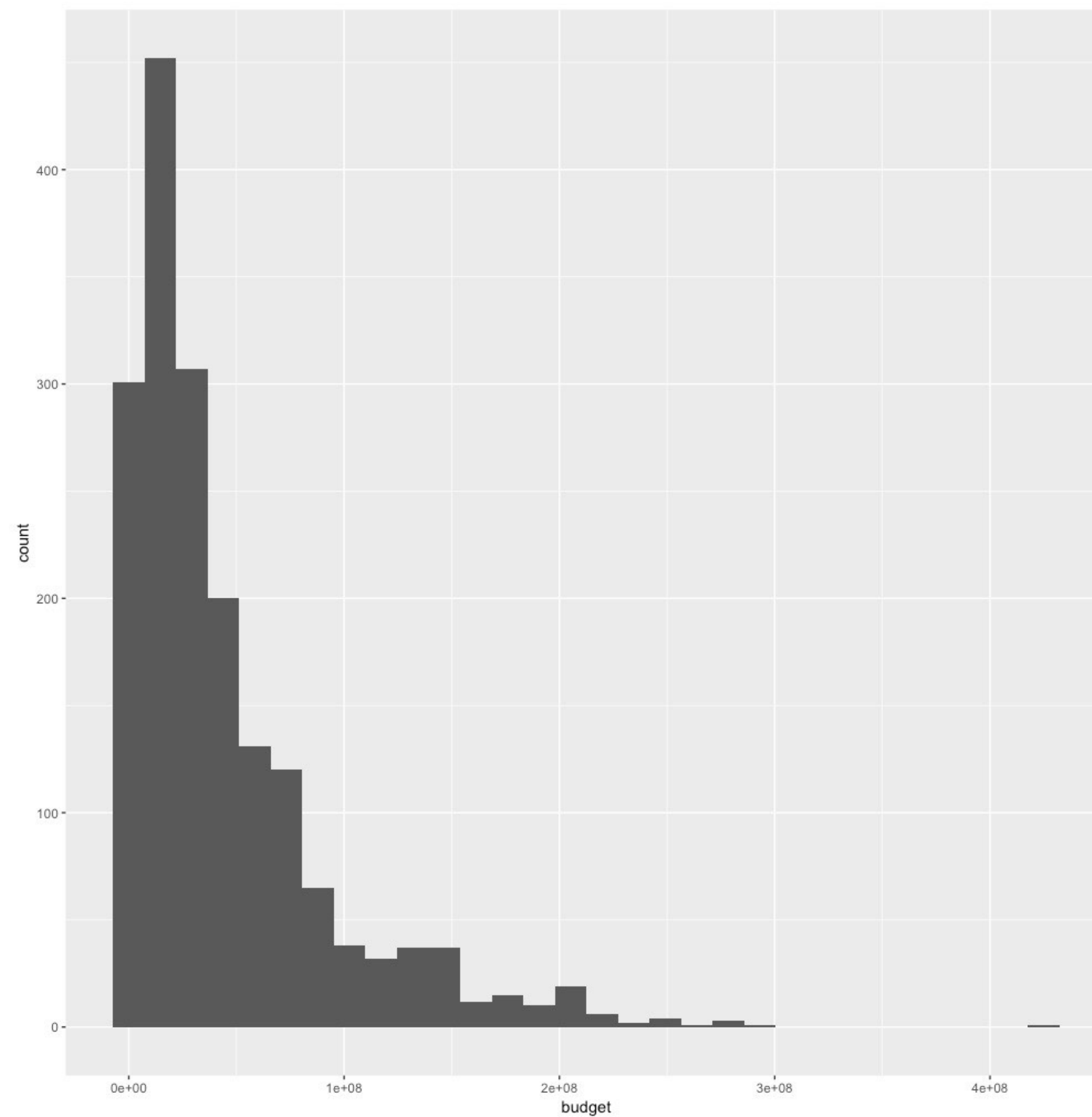
```
ggplot(data = bechdel) +  
  geom_histogram(mapping = aes(x = budget))
```



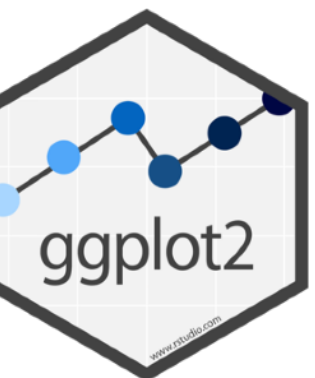
Your Turn 4

What would be a reasonable
`binwidth` for `budget`?

Try it out

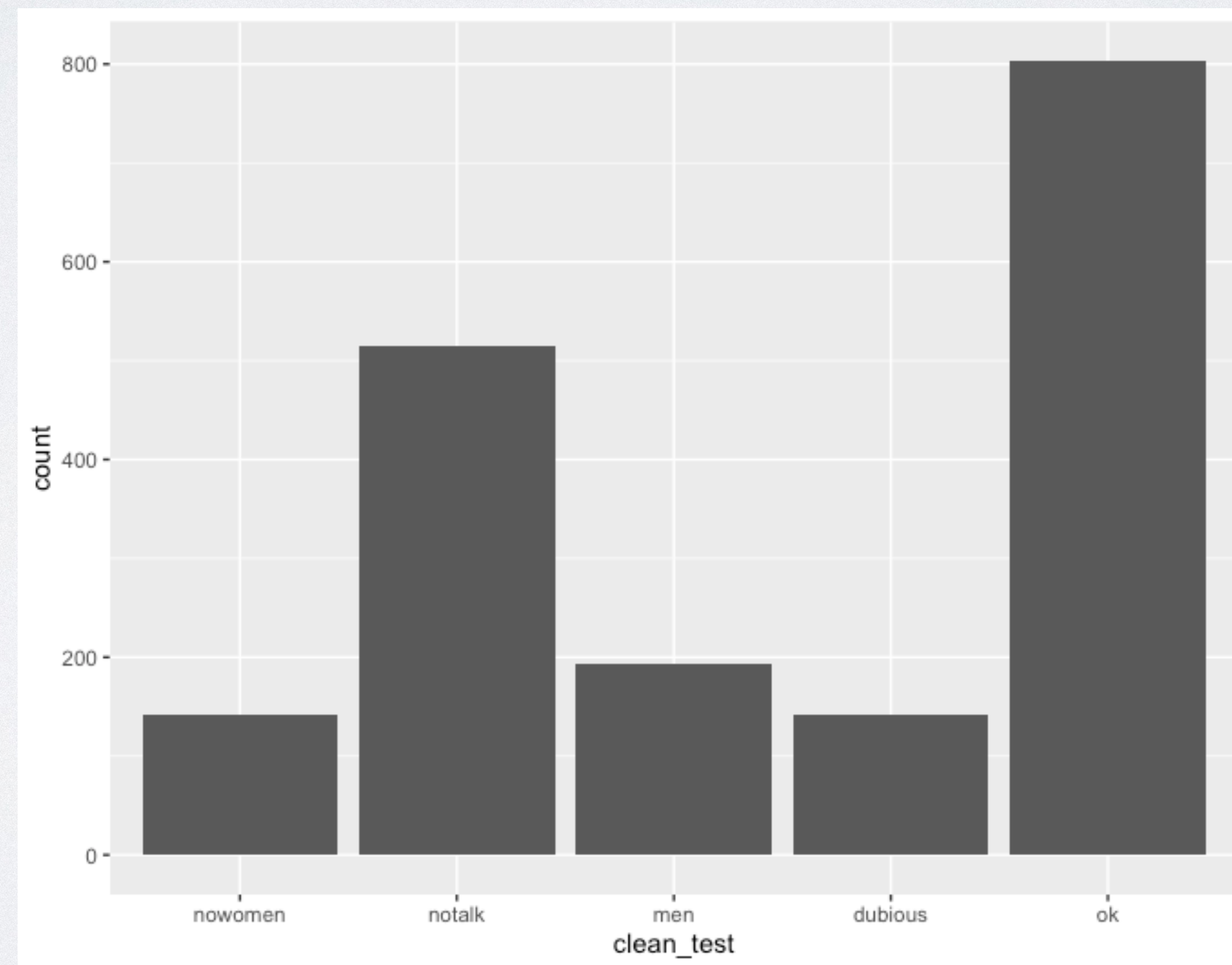


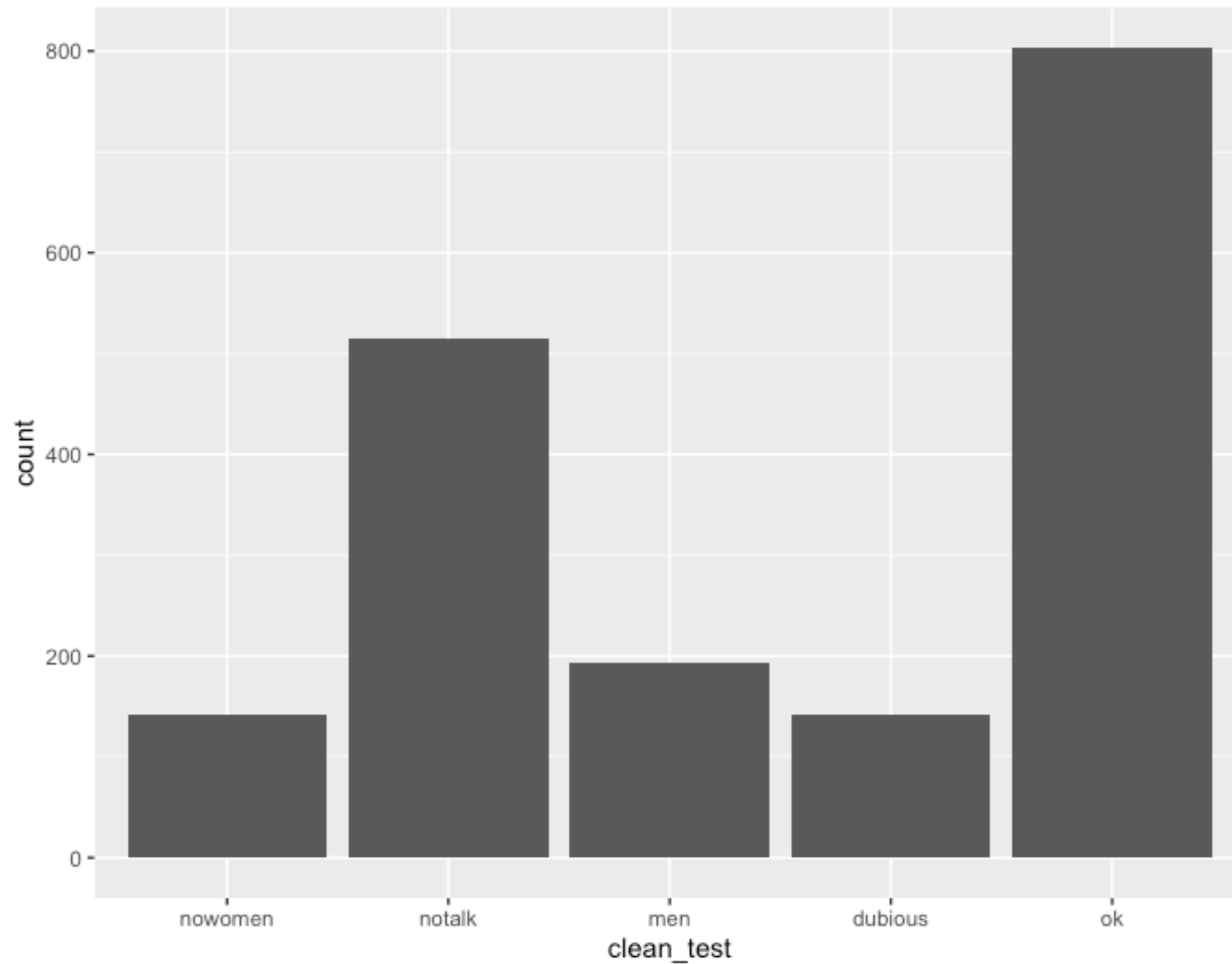
```
ggplot(data = bechdel) +  
  geom_histogram(mapping = aes(x = budget), binwidth=100000000)
```



Your Turn 5

With your partner, make a bar chart of **clean_test**.
Use the cheatsheet. Try your best guess.



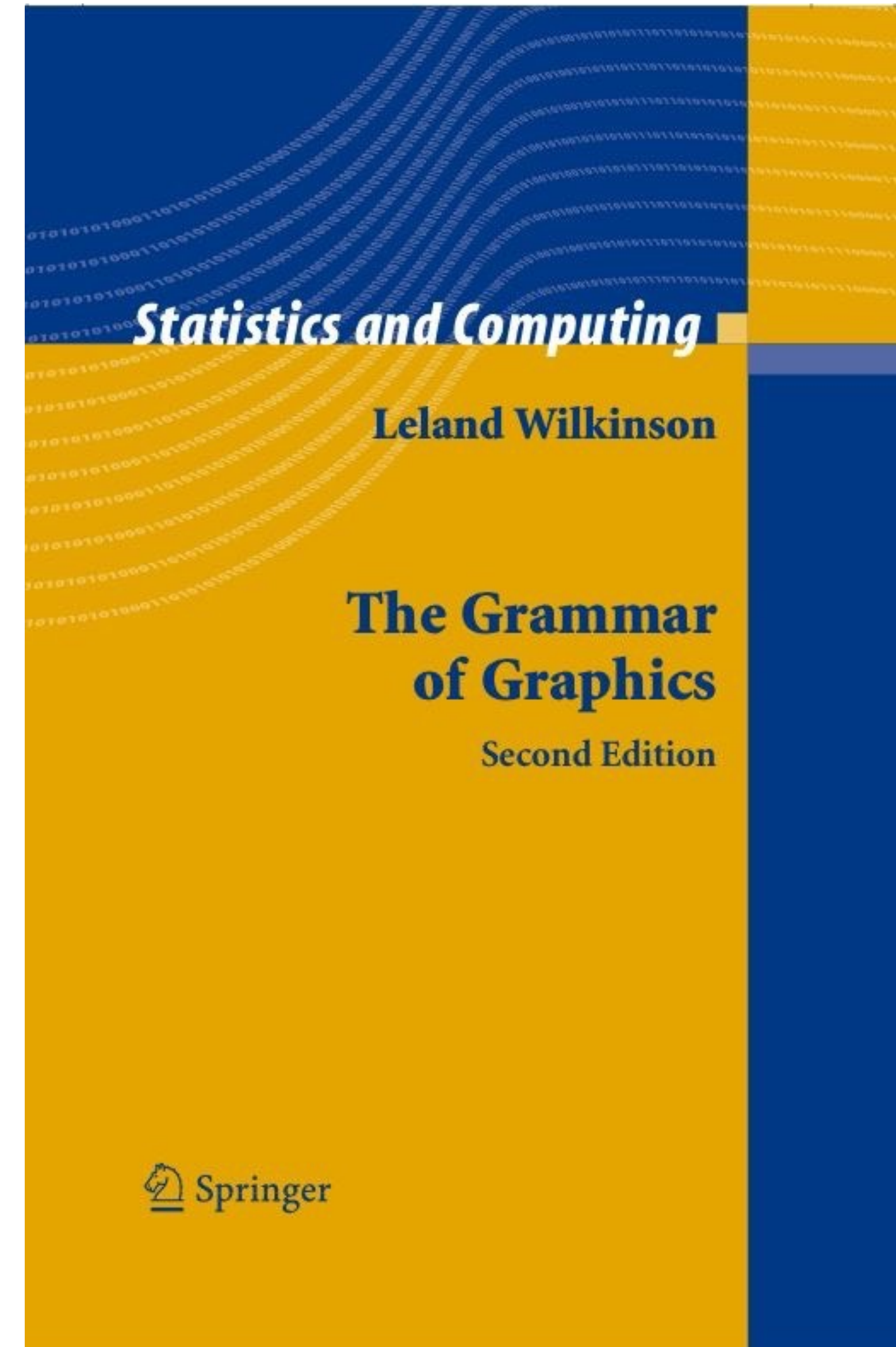


```
ggplot(data=bechdel) +  
  geom_bar(aes(x=clean_test))
```

faceting



The Grammar of Graphics

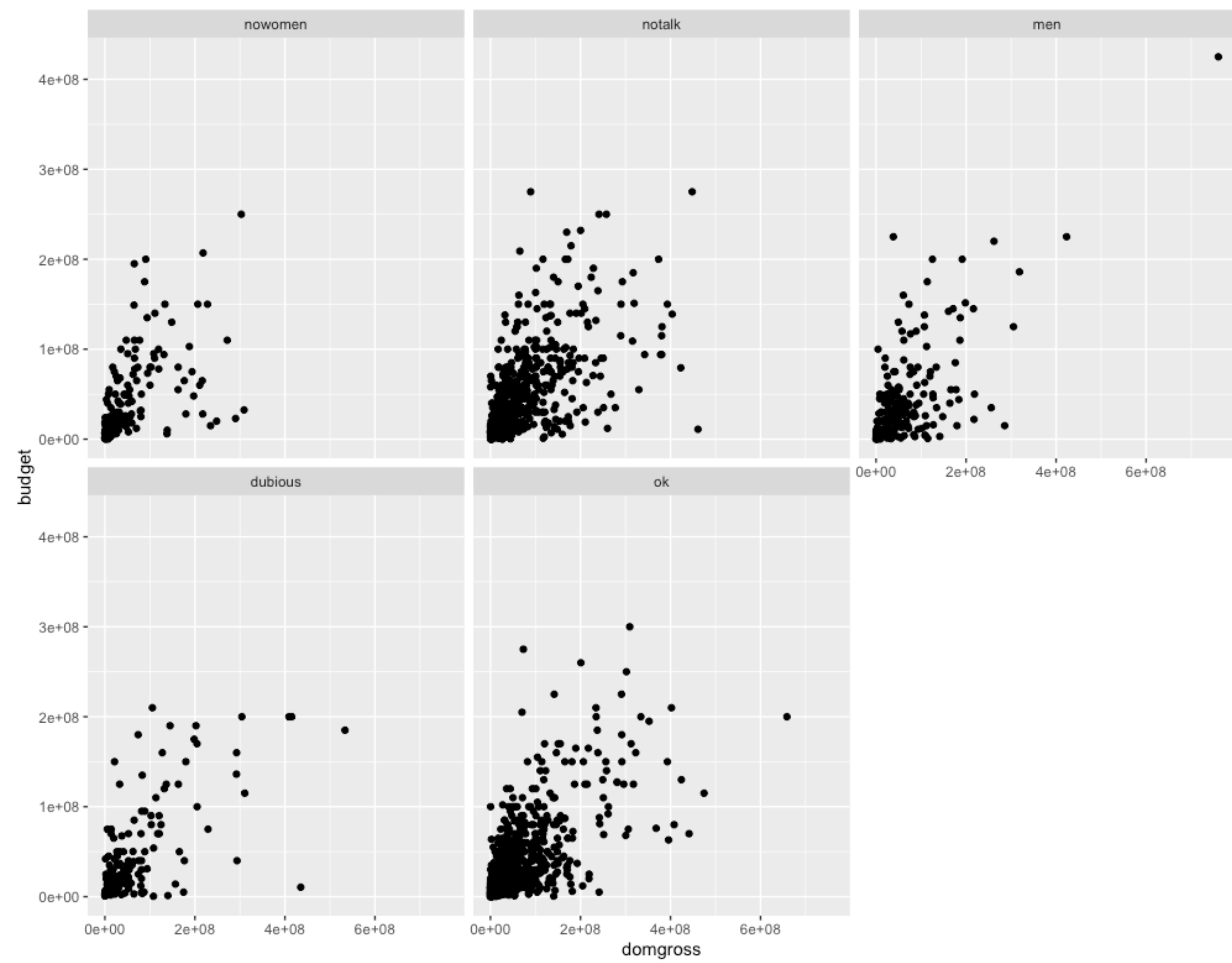


facets make the same plot many times

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>)) +  
  <FACET_FUNCTION>(vars(<VARIABLES>))
```

How are these plots similar?

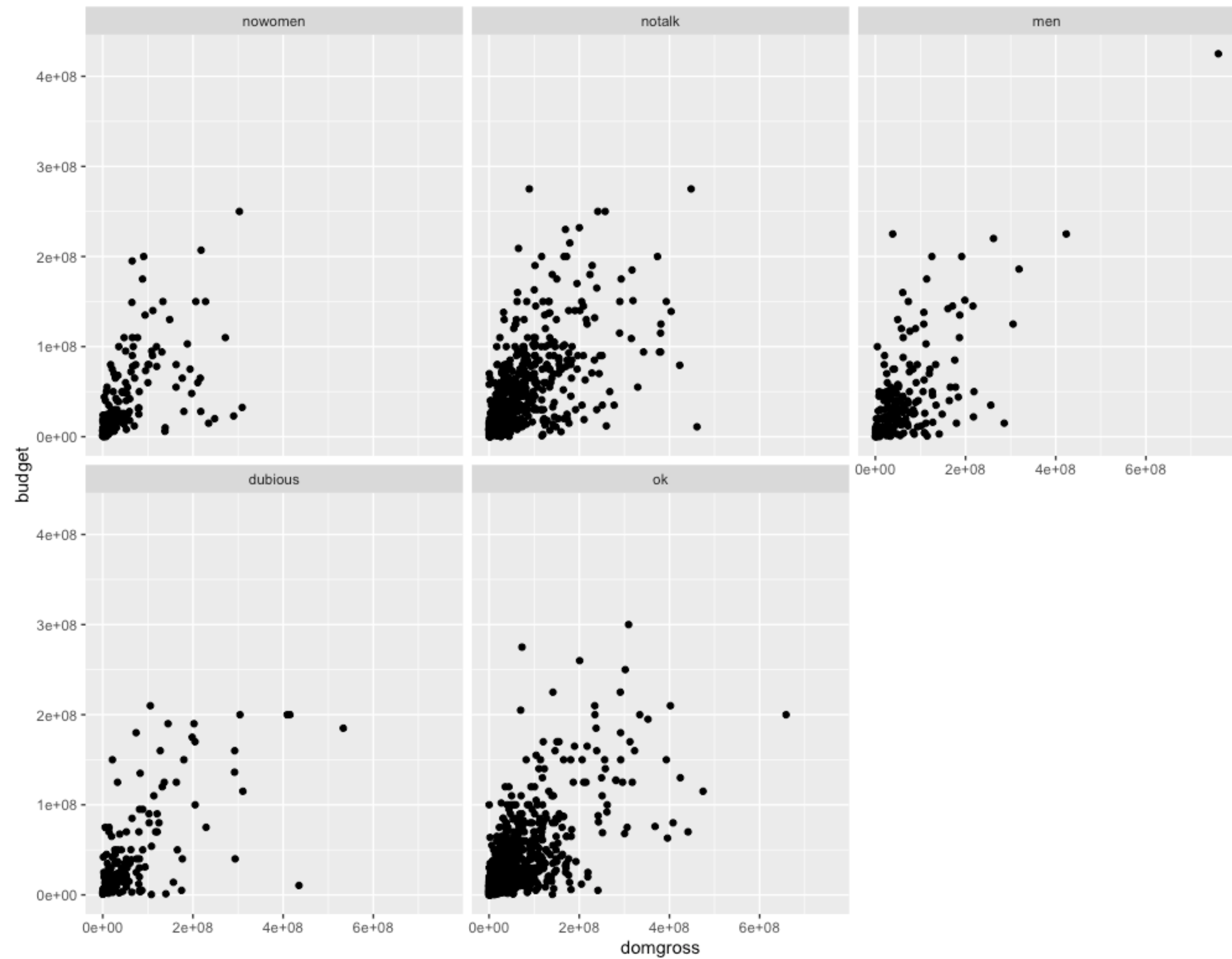
How are these plots different?



Your Turn 6

Run this code in your document to make a graph.

```
ggplot(data = bechdel) +  
  geom_point(aes(x = domgross, y = budget)) +  
  facet_wrap(vars(clean_test))
```

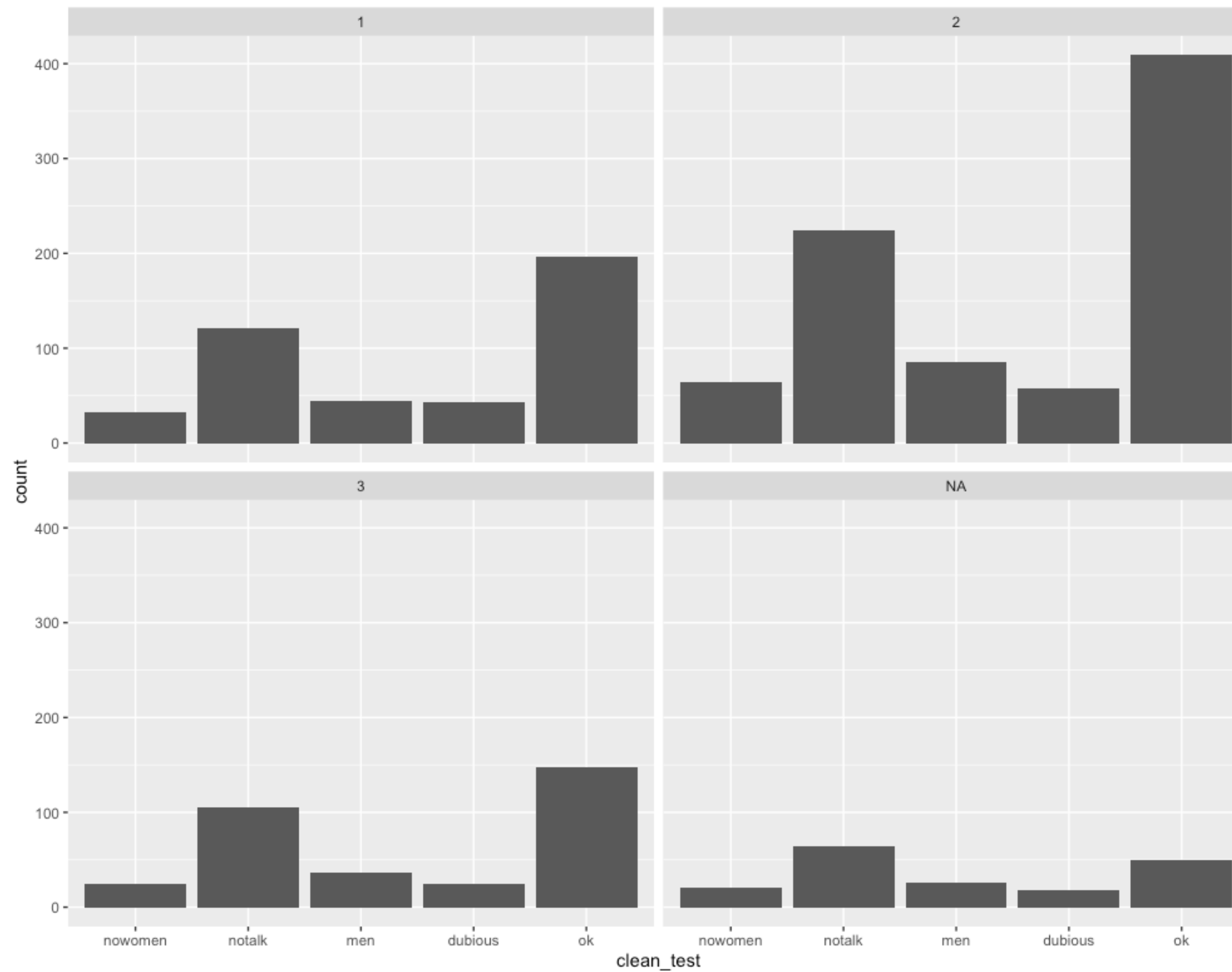


Your Turn 7

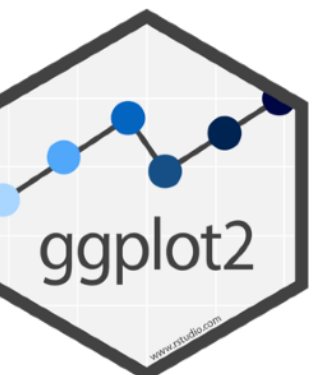
recall

```
ggplot(data = bechdel) +  
  geom_point(aes(x = domgross, y = budget)) +  
  facet_wrap(vars(clean_test))
```

With your partner, brainstorm how you could use faceting to show the relationship between two categorical variables



```
ggplot(data = bechdel) +  
  geom_bar(aes(x = clean_test)) +  
  facet_wrap(vars(decade_code))
```



Rendering



Rendering a Quarto document

The easiest way to save all your work (including graphs) is to include it in a Quarto document.
While you're working, you can render your document by clicking Render

The screenshot shows the RStudio interface with a Quarto document open. The document title is '01_ggplot2_intro.qmd'. The 'Render' button in the top toolbar is circled in red. The document content includes a dashed line, a section header '# Take aways', and a ggplot2 code template. The console shows R startup messages. The viewer pane displays a histogram of the 'count' variable from the 'bechdel' dataset.

```
108 as-is. Feel free to try out those options!
109
110 -----
111
112 # Take aways
113
114 You can use this code template to make thousands of graphs with ggplot2.
115
116 ```{r eval = FALSE}
117 ggplot(data = <DATA>) +
118   <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
119 ```
120
121 <!-- This file by Amelia McNamara is licensed under a Creative Commons Attribution 4.0 International
122      License, adapted from the original work at https://github.com/rstudio/master-the-tidyverse by RStudio.
123      -->
```

96:1 # Your Turn 4: histogram binwidth

Console: R 4.2.1 · ~/Documents/STAT336/

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Environment: 312 MiB

Data: bechdel (1794 obs. of 15 variables)

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export

count

800

600

400

200

ggplot2

HTML preview

Now, you'll see a beautifully typeset version of what you've done!

```
01_ggplot2_intro.qmd x
← → | 📄 | 📁 | Render on Save | ABC 🔍 | ➡ Render ⚙️ | +c | ↑ | ↓ | ➡ Run | ↻
Source Visual Outline

as-is. Feel free to try out those options!
108
109
110 -----
111
112 # Take aways
113
114 You can use this code template to make thousands of graphs with ggplot2.
115
116 ```{r eval = FALSE}
117 ggplot(data = <DATA>) +
118   <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
119 ```
120
121 <!-- This file by Amelia McNamara is licensed under a Creative Commons Attribution 4.0 International
License, adapted from the original work at https://github.com/rstudio/master-the-tidyverse by RStudio.
-->
122

114:55 # Take aways ⌵ Quarto ⌵
```

```
Console Background Jobs x
R 4.2.1 · ~/Documents/STAT336/
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(fivethirtyeight)
> library(ggplot2)
> data(bechdel)
```

Environment History Git Tutorial

Files Plots Packages Help Viewer Presentation

✖ 📄 ✎ Edit 🔍 ↻

Intro to ggplot2

AUTHOR
Professor McNamara

R packages

To use R packages, they need to be installed (once per computer) and loaded (once per R session). If you haven't already, you can install all the packages I think we might need for the semester by running the command below. You can run it by either pressing the green triangle button (I call it the play button), using the Run menu up at the top of this document, or putting your cursor on the code and pressing Command+Return (Mac) or Ctrl+Enter (PC).

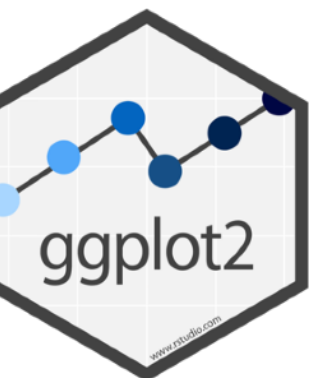
```
install.packages(c("tidyverse", "babynames", "br",
  "cowplot", "devtools", "drat",
  "fivethirtyeight", "formatR",
  "GGally", "ggforce", "ggraph",
  "ggridges", "graphlayouts", "
  "here", "hexbin", "interplot",
  "margins", "mgcv", "maps", "m
```



Sharing your work

The preview is something that pops up for you to look at as you work, but RStudio automatically creates a file you can share with others when you render a Quarto document. By default, it creates an HTML file, and the file's name corresponds to the name of your Quarto document.

You can email this HTML document to anyone you'd like to share your work with!



Sharing your work

The only trick is, images won't appear unless you have the option of “embed resources” set as true.

```
---  
title: "Intro to ggplot2"  
author: "Professor McNamara"  
format: html  
editor: visual  
execute:  
  echo: true  
embed-resources: true  
---
```

