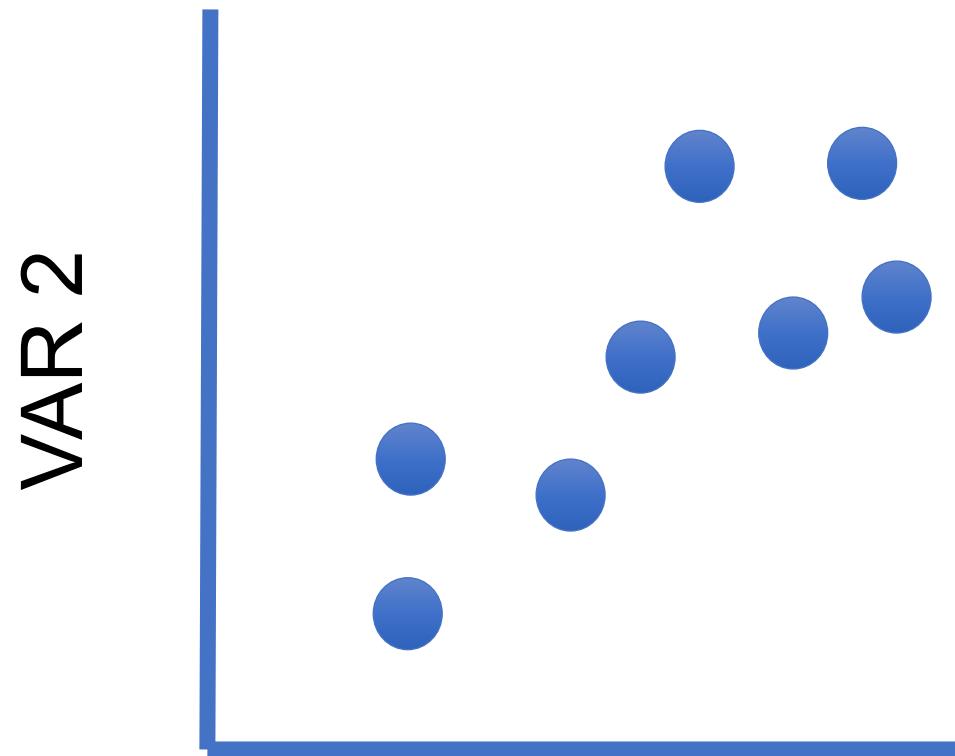


# ggplot2: A layered Grammar of Graphics

# What is this?:



# What is this?:

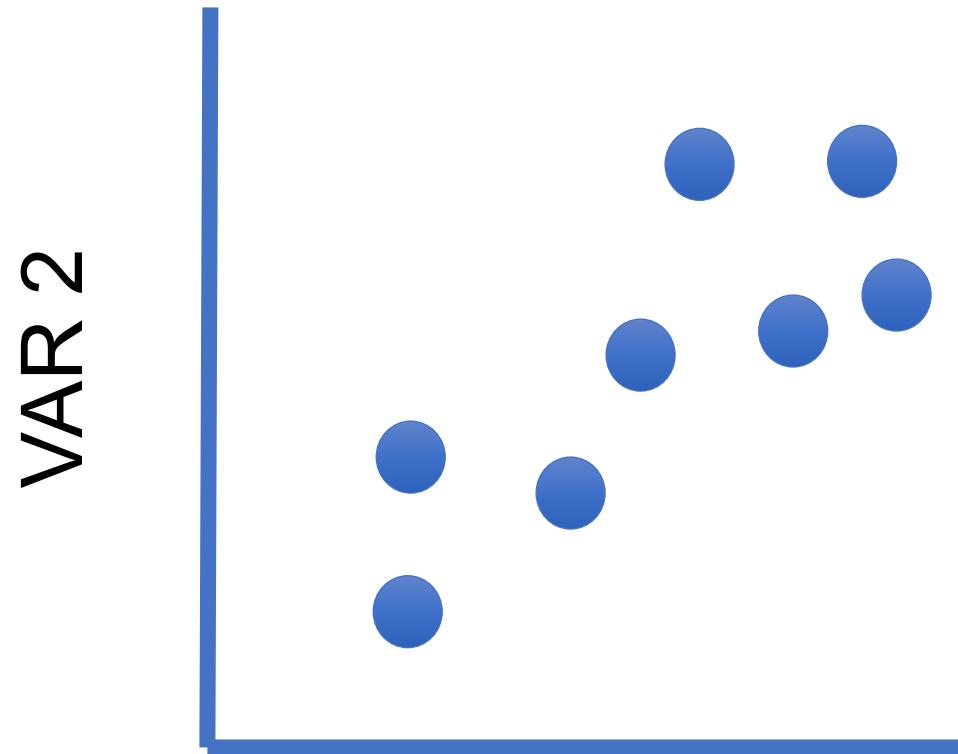
# What is this?: A coordinate system



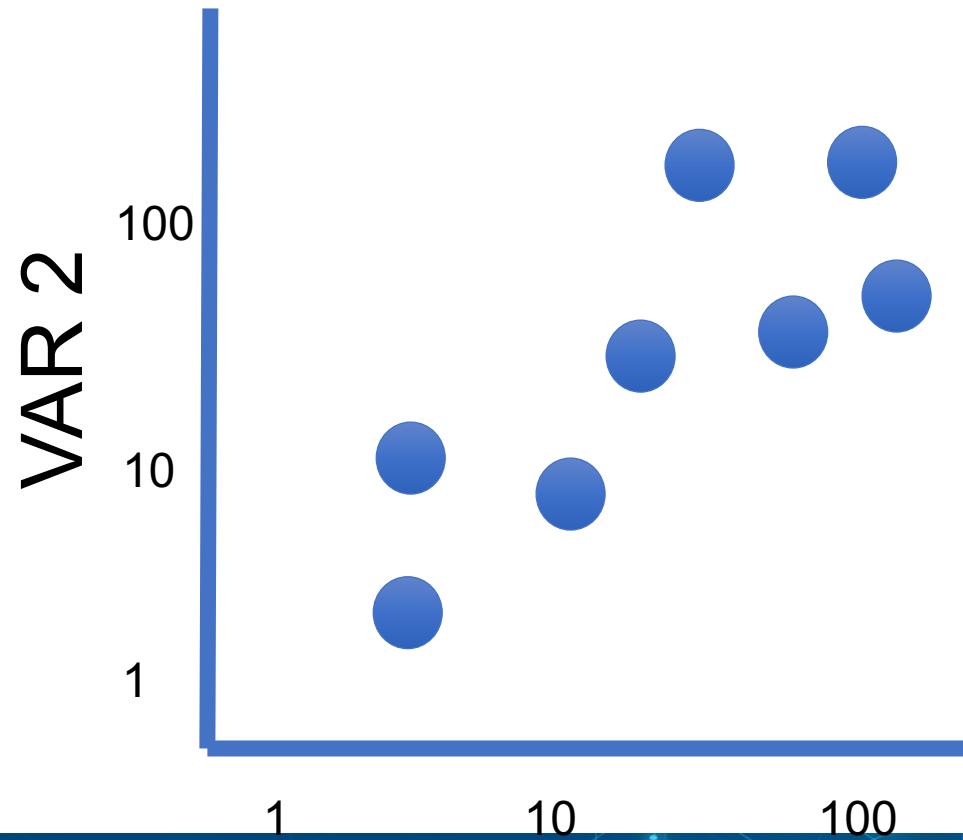
# What is this?: A coordinate system AND data and aesthetics



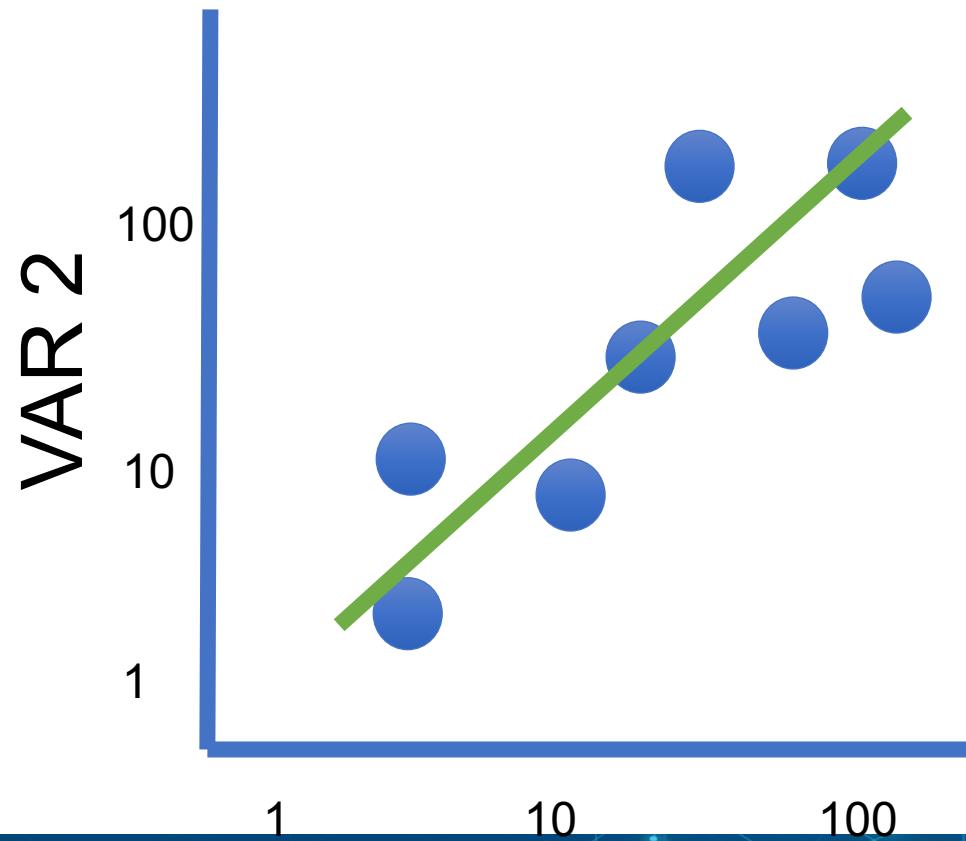
What is this?: A coordinate system, data and aesthetics AND geometric objects



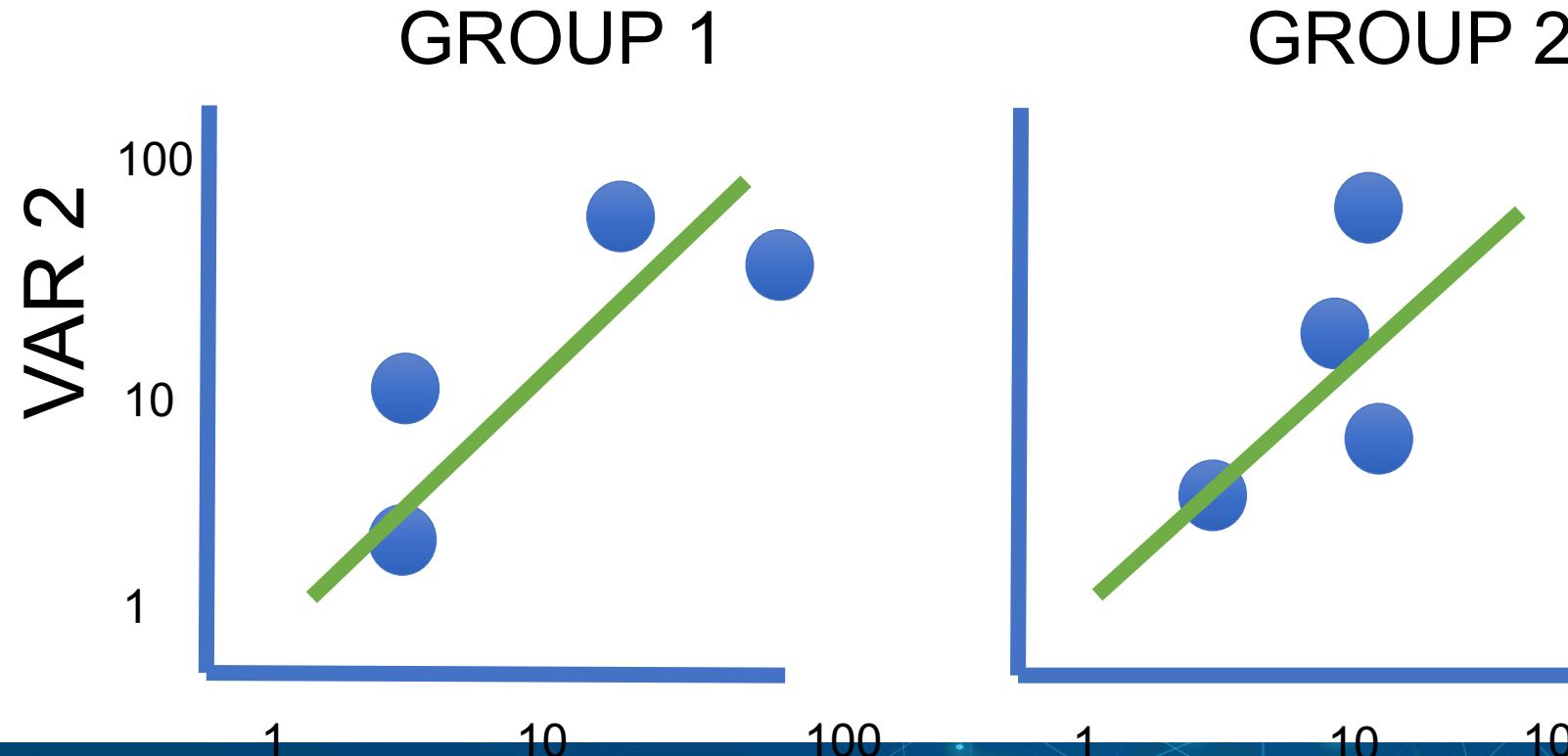
What is this?: A coordinate system data and aesthetics and geometric objects AND scale



What is this?: A coordinate system data and aesthetics and geometric objects and scale  
AND statistical transformations



What is this?: A coordinate system data and aesthetics and geometric objects and scale and statistical transformations AND Faceting



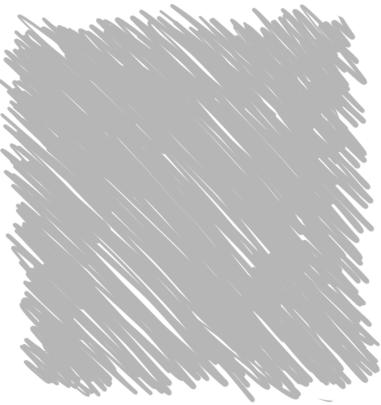
```
ggplot(data = mpg)
```



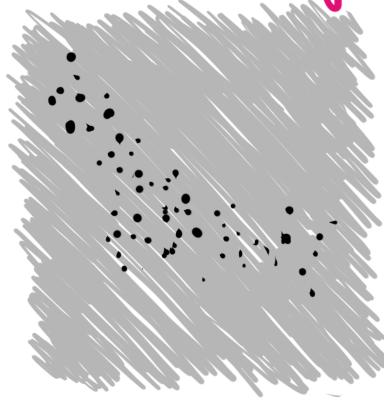
the university  
for the real world

Images from the amazing Assistant Professor Danielle Navarro

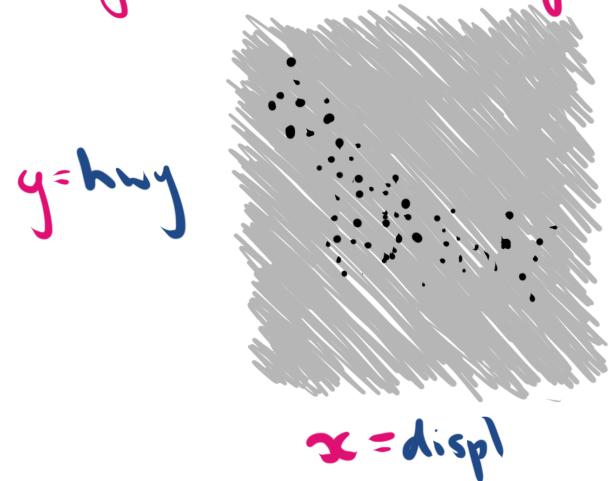
`ggplot(data = mpg)`



```
ggplot(data = mpg) +  
  geom_point(mapping = blah)
```

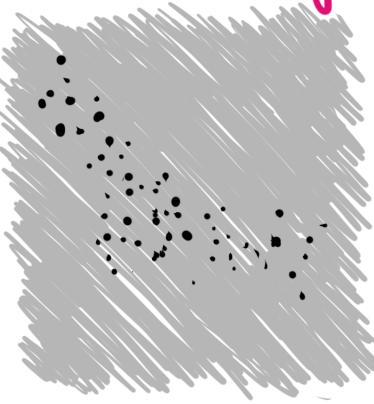


```
ggplot(data = mpg) +  
  geom_point(mapping = blah)
```



```
ggplot(data = mpg) +  
  geom_point(mapping = blab)
```

$y = \text{hwy}$   
miles per gallon  
for highway  
driving



$x = \text{displ}$   
engine displacement  
in litres.

# The real code

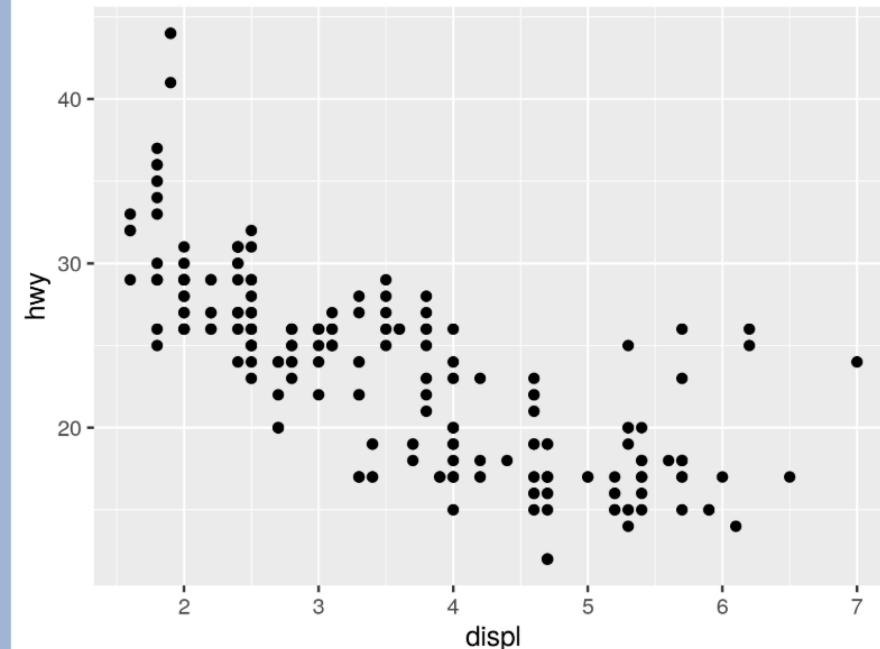
T

```
library(tidyverse)
ggplot(data = mpg) +
  geom_point(
    mapping = aes(
      x = displ,
      y = hwy
    )
  )
```

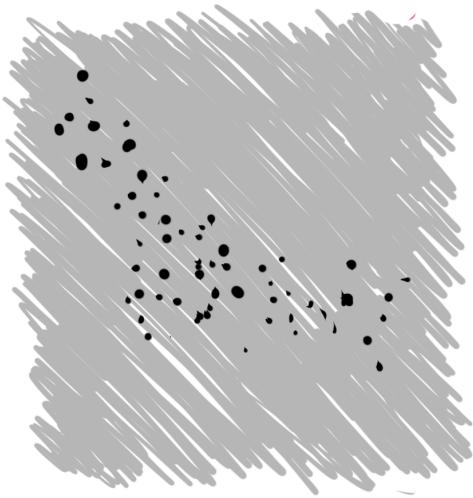
# The real code

```
library(tidyverse)
ggplot(data = mpg) +
  geom_point(
    mapping = aes(
      x = displ,
      y = hwy
    )
  )
```

# The real output



this plot uses mpg as the data



this plot uses mpg as the data

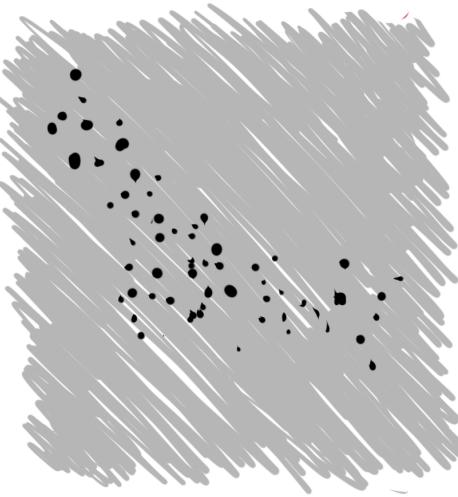


it contains  
variables

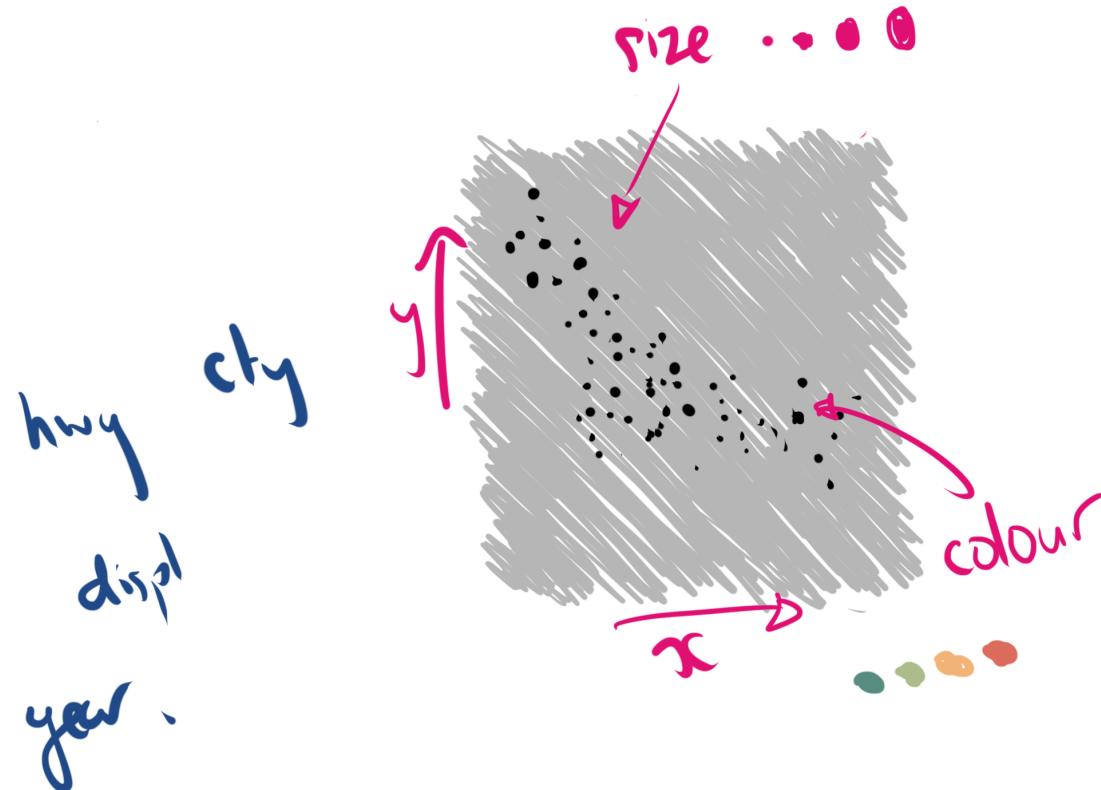
hwy cty

displ

year

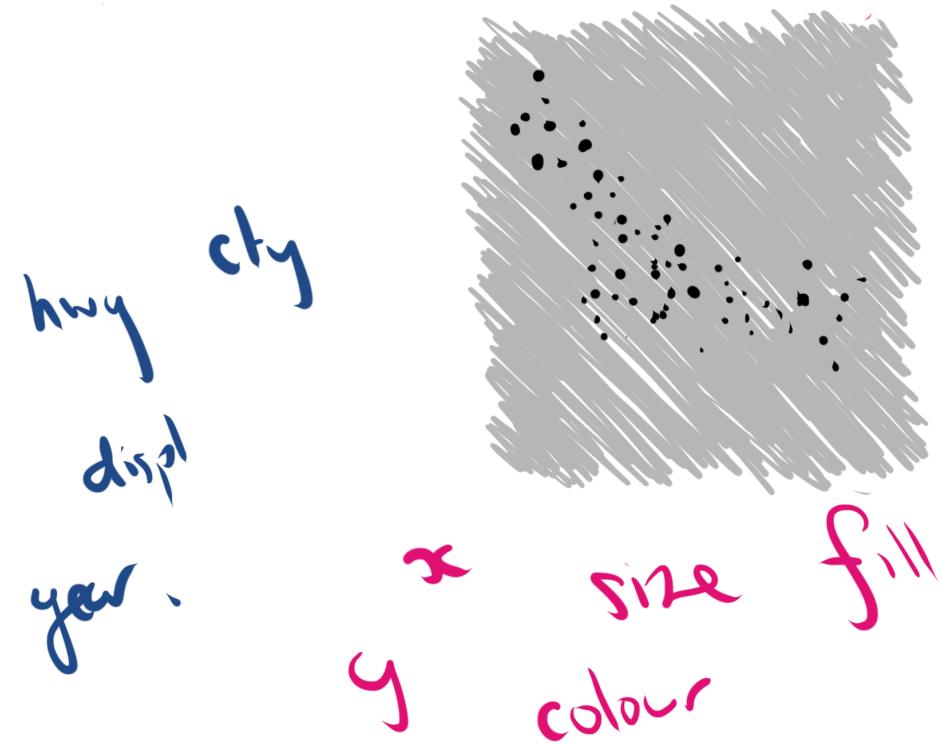


plot features are called  
“aesthetics”

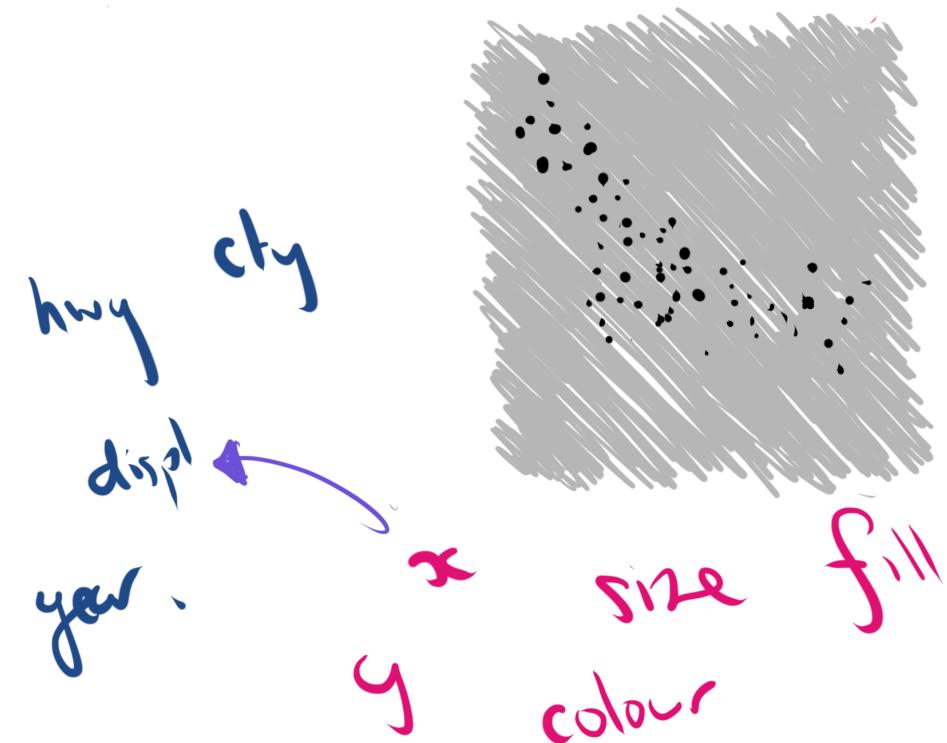


`aes(x = displ, y = hwy)`

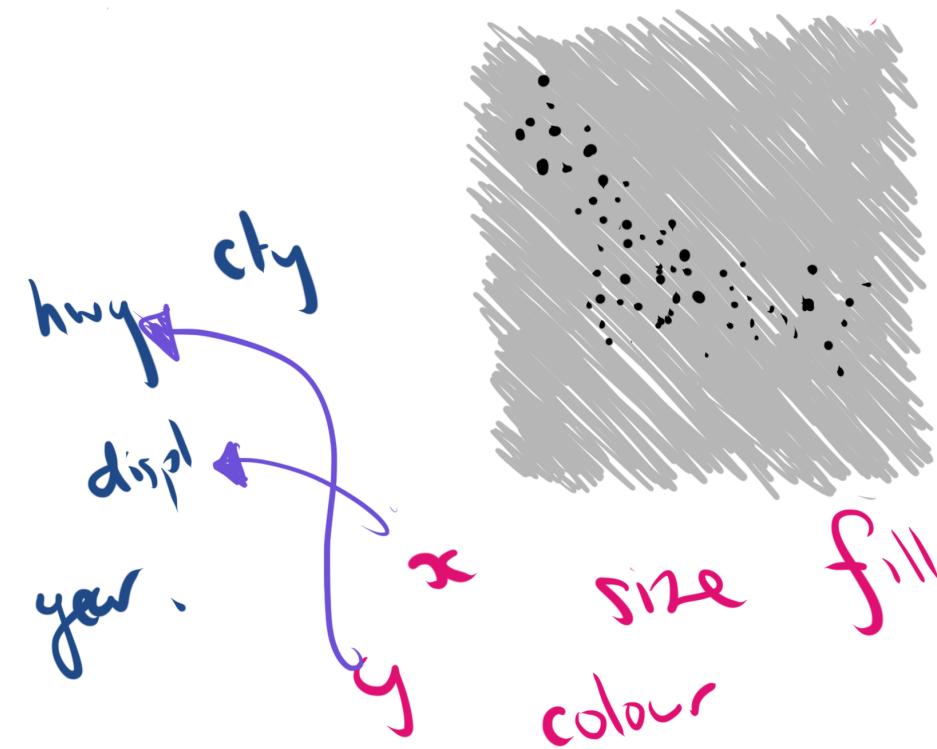
describes an "aesthetic mapping"



`aes(x = displ, y = hwy)`



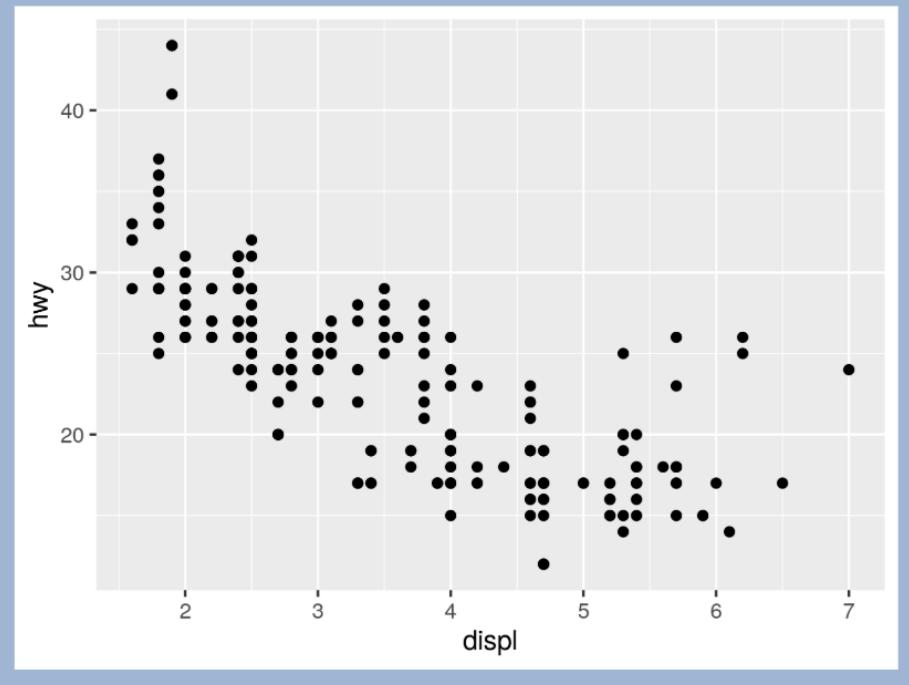
`aes(x = displ, y = hwy)`



# The real code

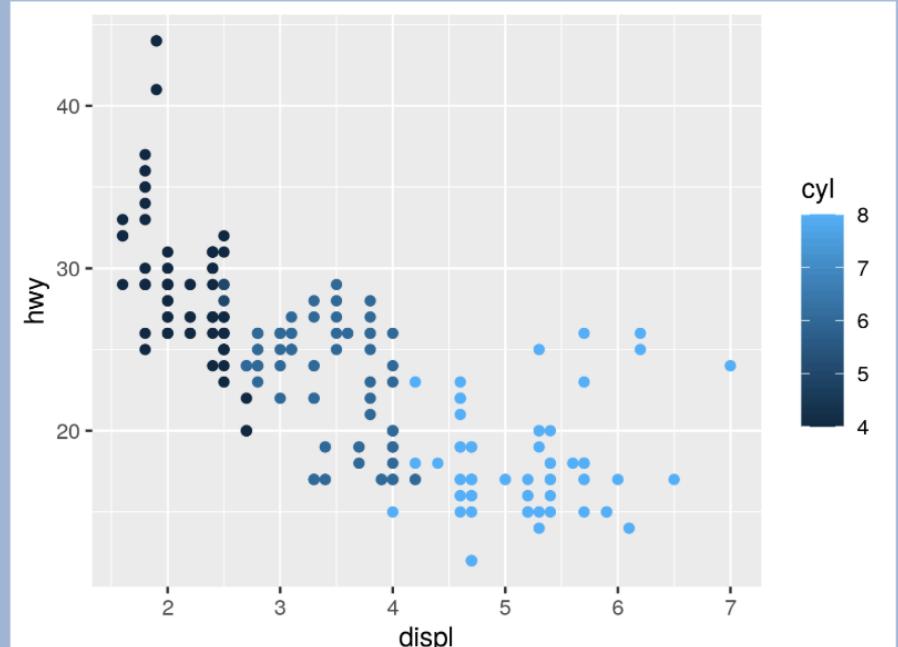
```
ggplot(data = mpg) +  
  geom_point(  
    mapping = aes(  
      x = displ,  
      y = hwy  
    )  
  )
```

# The real output



# Adding colour

```
ggplot(data = mpg) +  
  geom_point(  
    mapping = aes(  
      x = displ,  
      y = hwy,  
      color = cyl  
    )  
  )
```



# ggplot structure

```
ggplot(data = <DATA>) +
```

# ggplot structure

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION> (
```

# ggplot structure

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

Required

# ggplot structure

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

# ggplot structure

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(  
    mapping = aes(<MAPPINGS>),  
    stat = <STAT>,  
    position = <POSITION>  
) +
```

# ggplot structure

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(  
    mapping = aes(<MAPPINGS>),  
    stat = <STAT>,  
    position = <POSITION>  
  ) +  
  <COORDINATE_FUNCTION> +  
  <FACET_FUNCTION> +  
  <SCALE_FUNCTION> +  
  <THEME_FUNCTION>
```

Required

Not required,  
sensible  
defaults  
supplied

# Every ggplot2 call is (basically) the same:

First, tell it you're using the ggplot function:

```
ggplot(
```

# Every ggplot2 call is (basically) the same:

Then, introduce your dataset, and the parts you care about:

```
ggplot(data = dataset, aes(x=variable1, y=variable2))
```

# Every ggplot2 call is (basically) the same:

Then, tell it what kind of plot you want

```
ggplot(dataset, aes(x=variable1, y=variable2)  
+ geom_[plottype]()
```

# Every ggplot2 call is (basically) the same:

Add your trimmings: labels, titles, colours, etc

```
ggplot(dataset, aes(x=variable1, y=variable2)
```

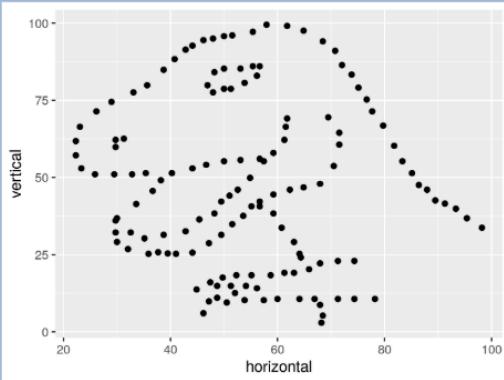
```
+ geom_[plottype]()
```

```
+ ggtile(your_title)
```

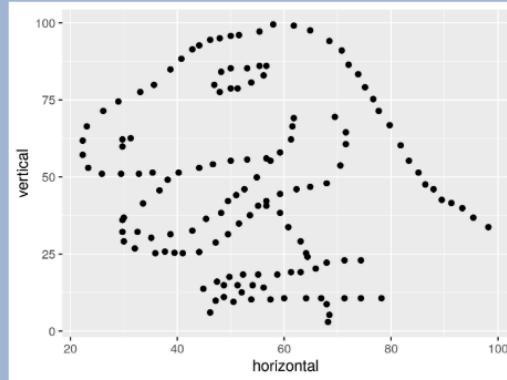
```
+ ylab(your_ylabel)
```

```
+ xlab(your_xlabel))
```

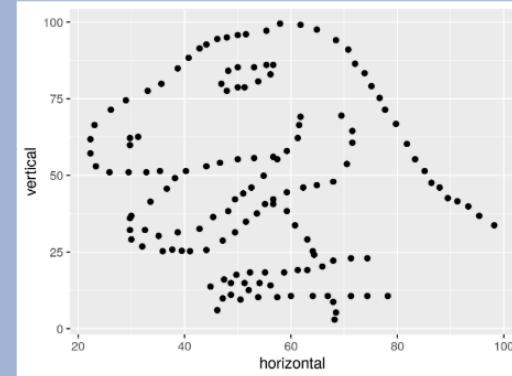
```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(
    mapping = aes(
      x = horizontal,
      y = vertical
    )
  )
plot(picture)
```



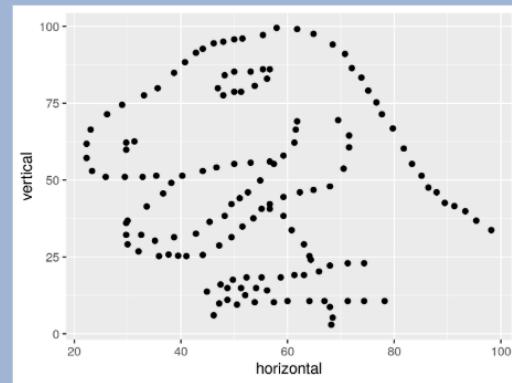
```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(
    mapping = aes(
      x = horizontal,
      y = vertical
    )
  )
plot(picture)
```



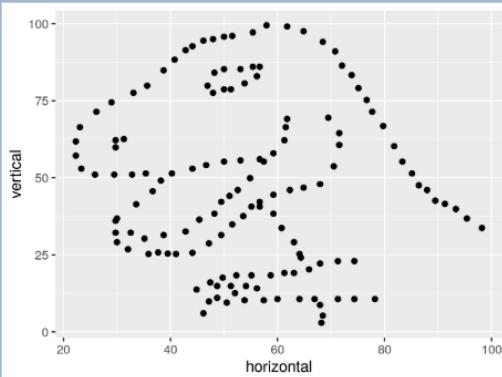
```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(
    aes(
      x = horizontal,
      y = vertical
    )
  )
plot(picture)
```



```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(
    aes(
      x = horizontal,
      y = vertical
    )
  )
plot(picture)
```



```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(
    aes(
      horizontal,
      vertical
    )
  )
plot(picture)
```



```
dino ← read_csv("data_dino.csv")
picture ← ggplot(dino) +
  geom_point(aes(horizontal, vertical))

plot(picture)
```

