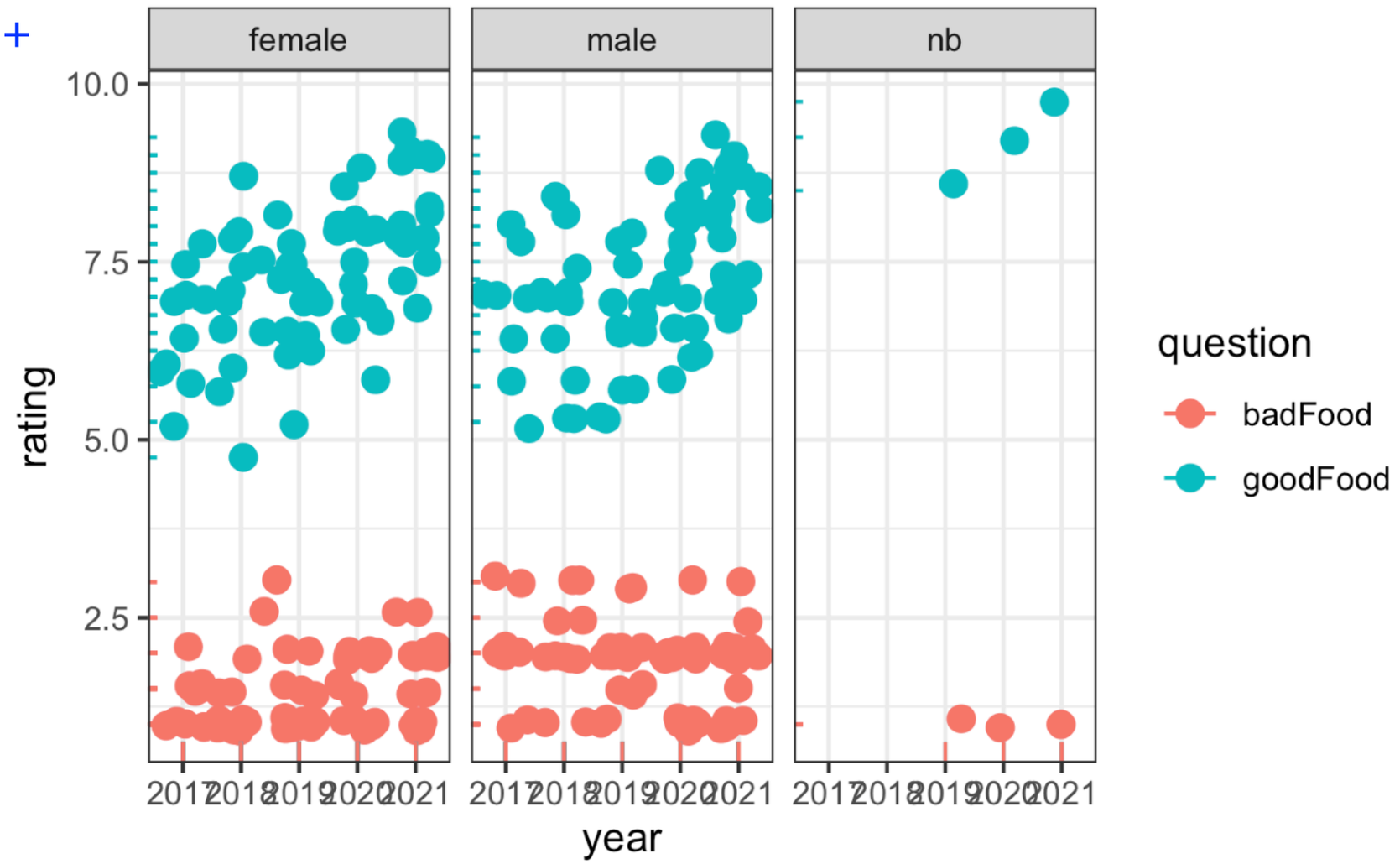
Visualisation: Geoms

Research Methods for Human Inquiry

Andrew Perfors

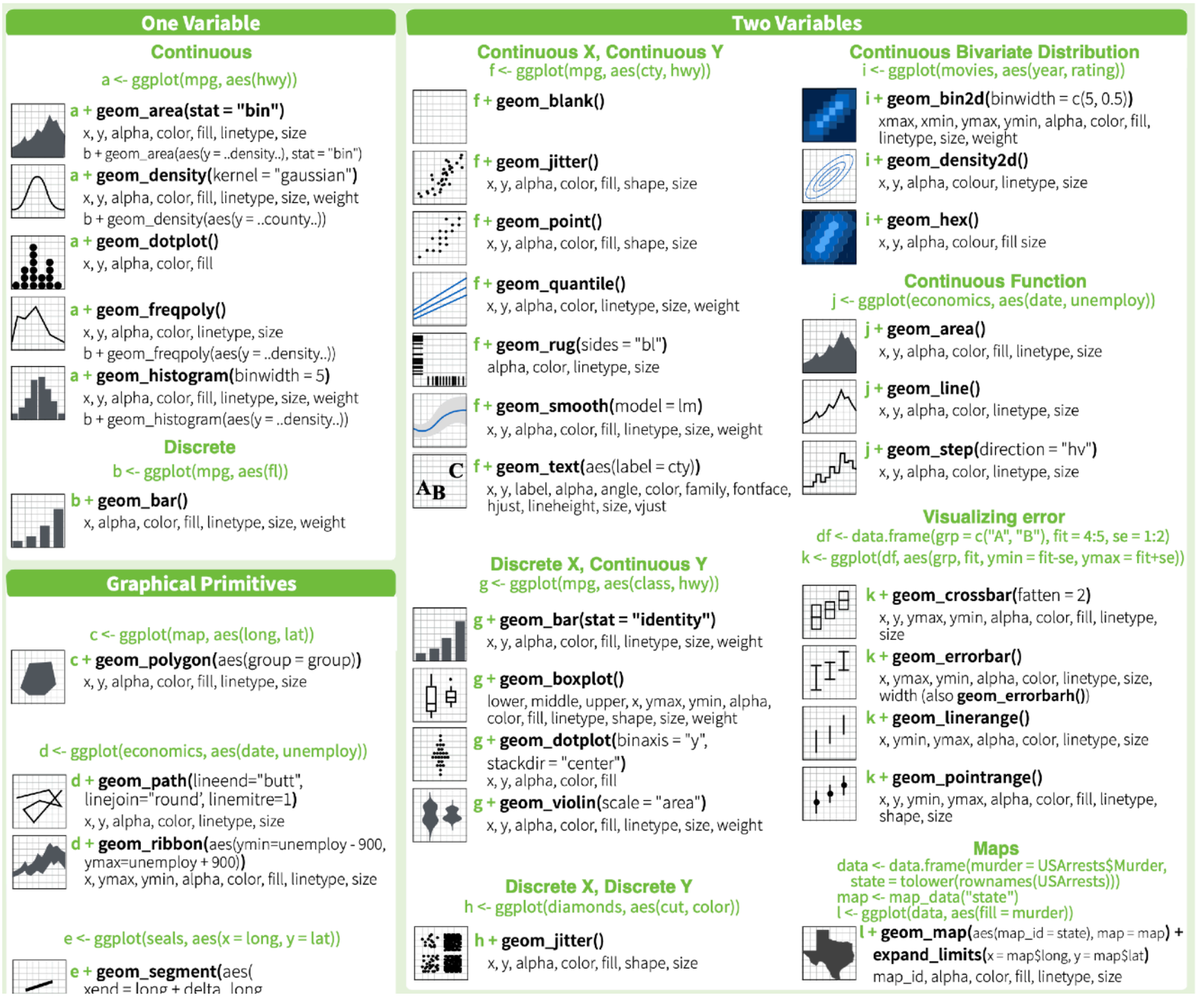
We've already seen some geoms

```
dl %>%
  ggplot(mapping = aes(x = year, y = rating, colour=question)) +
  geom_jitter(size=3) +
  geom_rug() +
  facet_wrap(~gender) +
  theme_bw()
```



Geoms define the look of your data

There are lots of them!



One Variable

Continuous

a <- ggplot(mpg, aes(hwy))

a + geom_area(stat = "bin")
x, y, alpha, color, fill, linetype, size
b + geom_area(aes(y = ..density..), stat = "bin")

a + geom_density(kernel = "gaussian"**)**
x, y, alpha, color, fill, linetype, size, weight
b + geom_density(aes(y = ..county..))

a + geom_dotplot()
x, y, alpha, color, fill

a + geom_freqpoly()
x, y, alpha, color, linetype, size
b + geom_freqpoly(aes(y = ..density..))

a + geom_histogram(binwidth = 5**)**
x, y, alpha, color, fill, linetype, size, weight
b + geom_histogram(aes(y = ..density..))

Discrete

b <- ggplot(mpg, aes(fl))

b + geom_bar()
x, alpha, color, fill, linetype, size, weight

Graphical Primitives

c <- ggplot(map, aes(long, lat))

c + geom_polygon(aes(group = group)**)**
x, y, alpha, color, fill, linetype, size

d <- ggplot(economics, aes(date, unemploy))

d + geom_path(lineend="butt", linejoin="round', linemitre=1**)**
x, y, alpha, color, linetype, size

d + geom_ribbon(aes(ymin=unemploy - 900, ymax=unemploy + 900)**)**
x, ymax, ymin, alpha, color, fill, linetype, size

e <- ggplot(seals, aes(x = long, y = lat))

e + geom_segment(aes(
xend = long + delta_long

Two Variables

Continuous X, Continuous Y

f <- ggplot(mpg, aes(cty, hwy))

f + geom_blank()

f + geom_jitter()
x, y, alpha, color, fill, shape, size

f + geom_point()
x, y, alpha, color, fill, shape, size

f + geom_quantile()
x, y, alpha, color, linetype, size, weight

f + geom_rug(sides = "bl"**)**
alpha, color, linetype, size

f + geom_smooth(model = lm**)**
x, y, alpha, color, fill, linetype, size, weight

f + geom_text(aes(label = cty)**)**
x, y, label, alpha, angle, color, family, fontface, hjust, lineheight, size, vjust

Discrete X, Continuous Y

g <- ggplot(mpg, aes(class, hwy))

g + geom_bar(stat = "identity")
x, y, alpha, color, fill, linetype, size, weight

g + geom_boxplot()
lower, middle, upper, x, ymax, ymin, alpha, color, fill, linetype, shape, size, weight

g + geom_dotplot(binaxis = "y", stackdir = "center"**)**
x, y, alpha, color, fill

g + geom_violin(scale = "area"**)**
x, y, alpha, color, fill, linetype, size, weight

Discrete X, Discrete Y

h <- ggplot(diamonds, aes(cut, color))

h + geom_jitter()
x, y, alpha, color, fill, shape, size

Continuous Bivariate Distribution

i <- ggplot(movies, aes(year, rating))

i + geom_bin2d(binwidth = c(5, 0.5)**)**
xmax, xmin, ymax, ymin, alpha, color, fill, linetype, size, weight

i + geom_density2d()
x, y, alpha, colour, linetype, size

i + geom_hex()
x, y, alpha, colour, fill size

Continuous Function

j <- ggplot(economics, aes(date, unemploy))

j + geom_area()
x, y, alpha, color, fill, linetype, size

j + geom_line()
x, y, alpha, color, linetype, size

j + geom_step(direction = "hv"**)**
x, y, alpha, color, linetype, size

Visualizing error

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

k + geom_crossbar(fatten = 2**)**
x, y, ymax, ymin, alpha, color, fill, linetype, size

k + geom_errorbar()
x, ymax, ymin, alpha, color, linetype, size, width (also **geom_errorbarh()**)

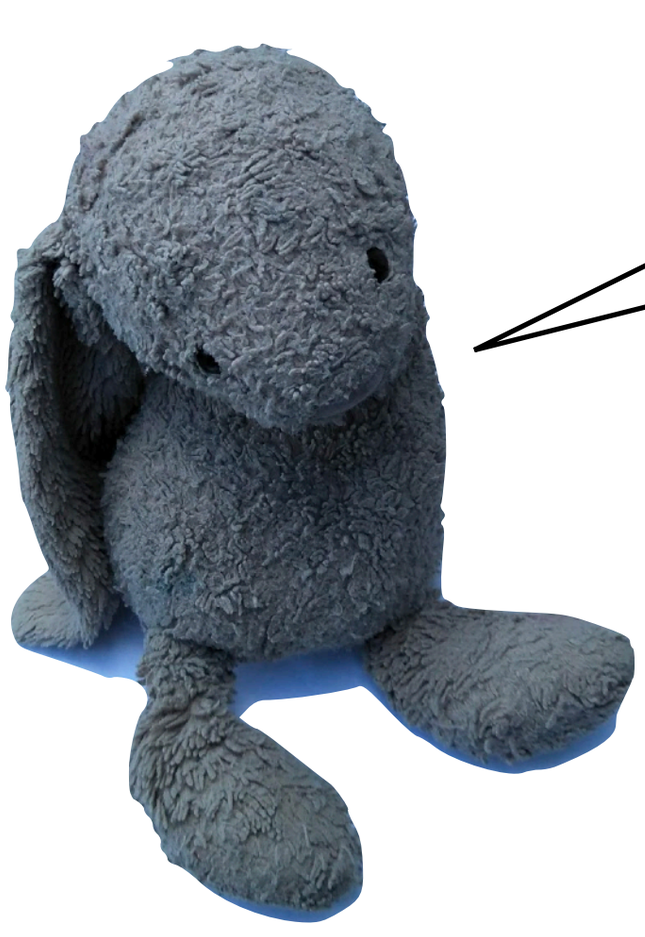
k + geom_linerange()
x, ymin, ymax, alpha, color, linetype, size

k + geom_pointrange()
x, y, ymin, ymax, alpha, color, fill, linetype, shape, size

Maps

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

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



OMG that is too much!!!!!

You don’t need to memorise them all!

Just understand basically how they work, and be able to look them up when you need to!

Different kinds based on the **number and type of variables** they take

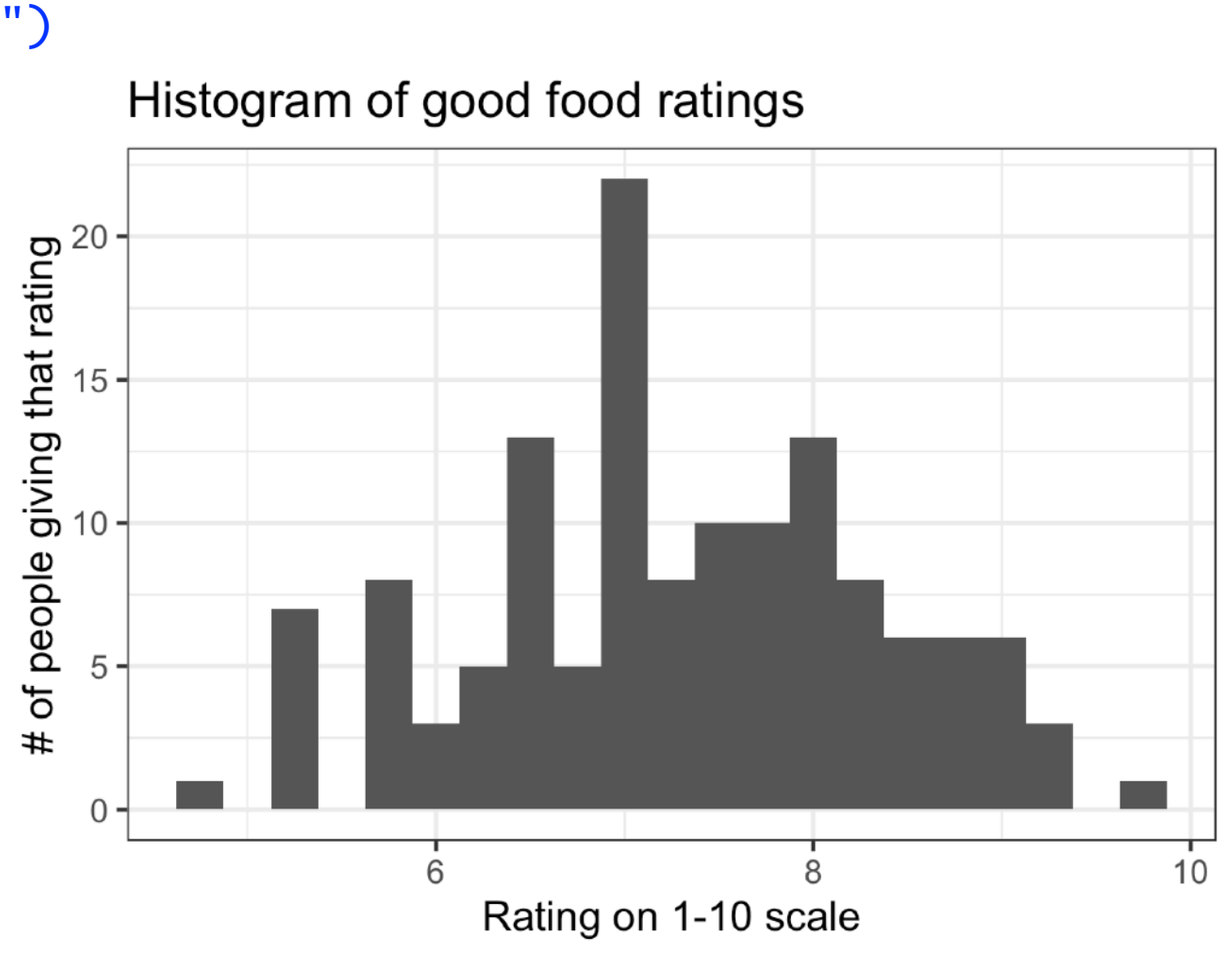
Continuous or discrete?

1 or 2?
(or more, but don’t worry about those right now)

One variable, continuous

Example: histogram

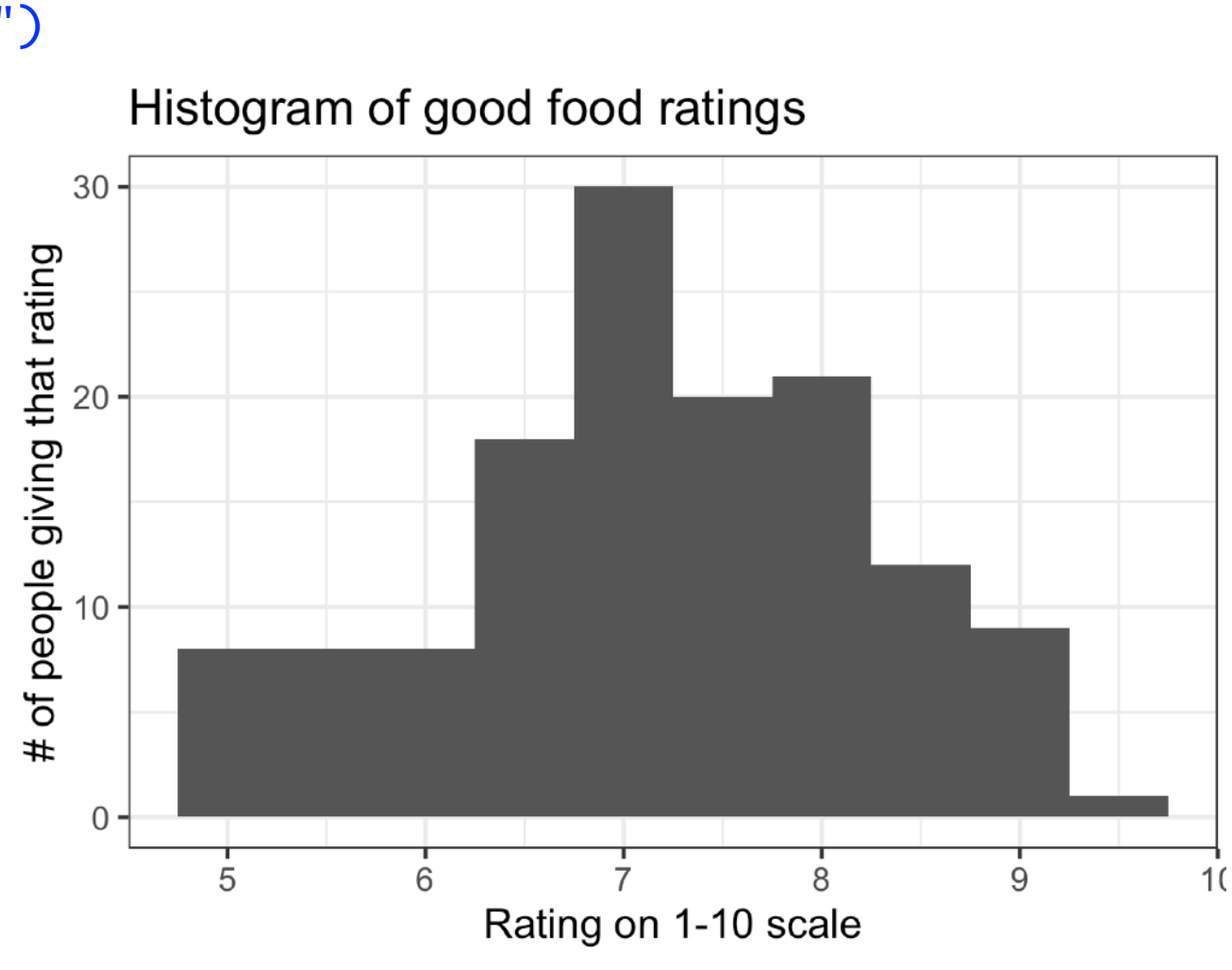
```
dnew %>%
  ggplot(mapping = aes(x = goodFood) ) +
  geom_histogram(binwidth=0.25) +
  theme_bw() +
  labs(title = "Histogram of good food ratings",
    x = "Rating on 1-10 scale",
    y = "# of people giving that rating")
```



One variable, continuous

Example: histogram

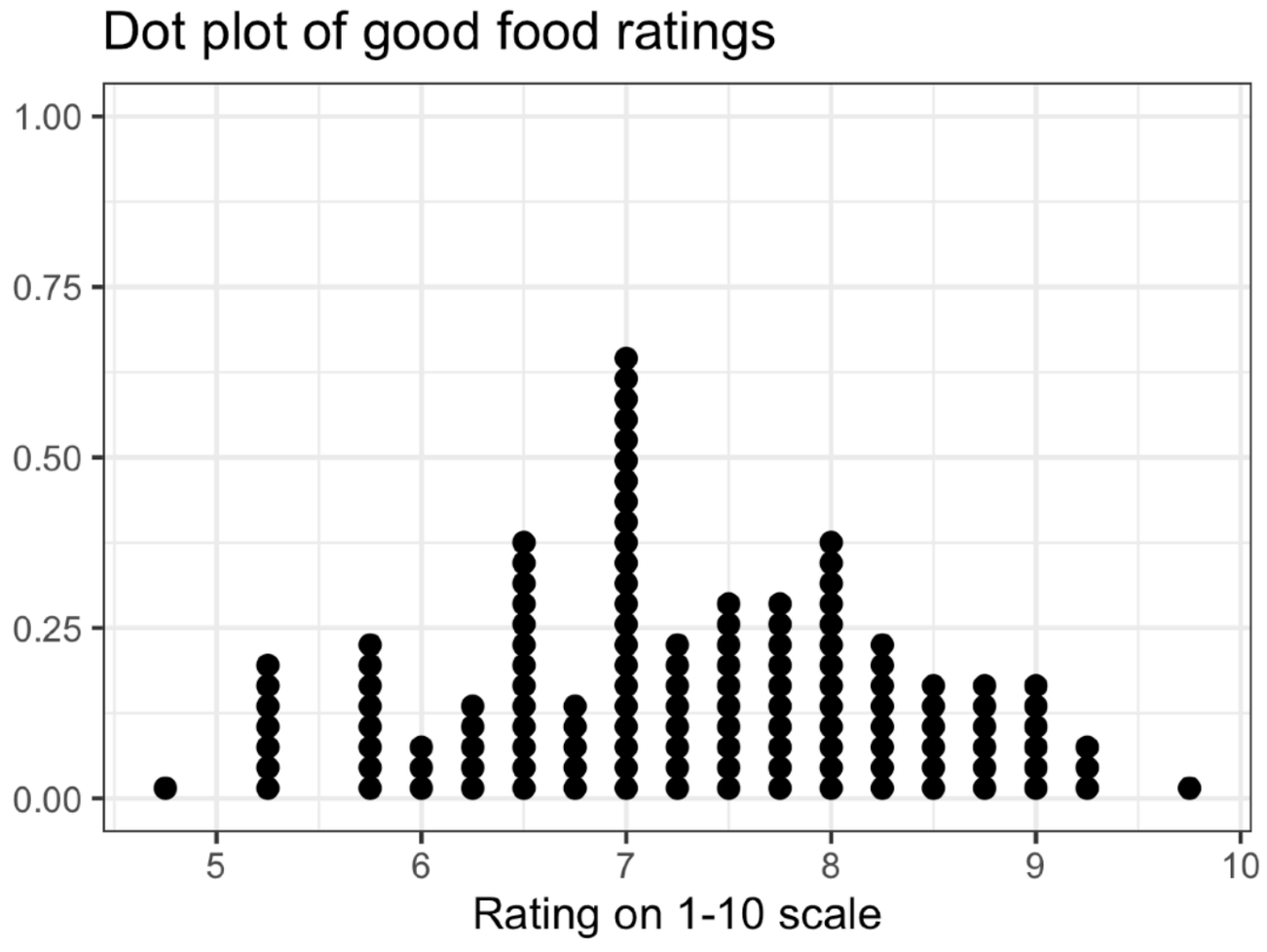
```
dnew %>%
  ggplot(mapping = aes(x = goodFood) ) +
  geom_histogram(binwidth=0.5) +
  theme_bw() +
  labs(title = "Histogram of good food ratings",
    x = "Rating on 1-10 scale",
    y = "# of people giving that rating")
```



One variable, continuous

Example: dot plot

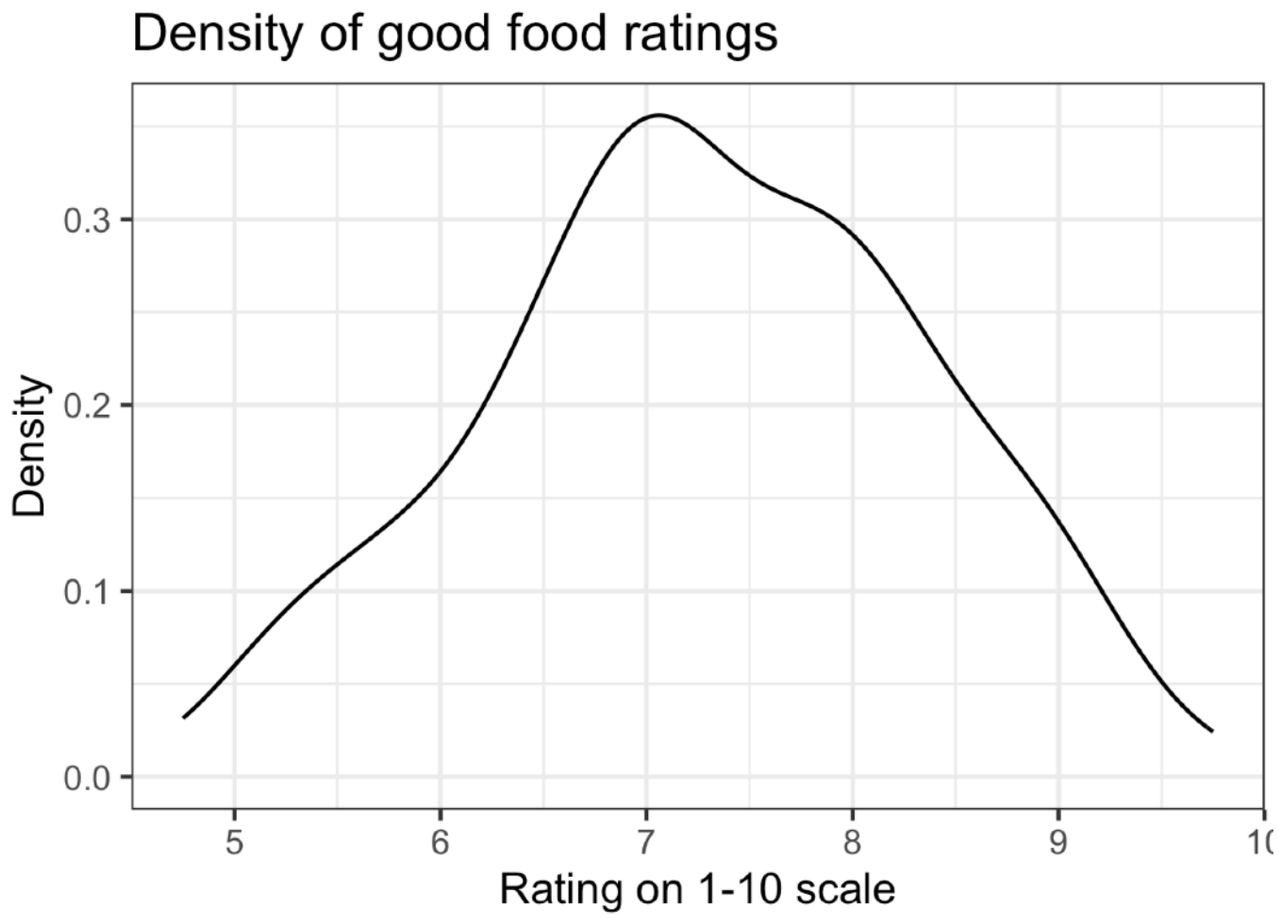
```
dnew %>%
  ggplot(mapping = aes(x = goodFood) ) +
  geom_dotplot(binwidth=0.1) +
  theme_bw() +
  labs(title = "Dot plot of good food ratings",
       x = "Rating on 1-10 scale",
       y = "")
```



One variable, continuous

Example: density (a kind of smoothed histogram)

```
dnew %>%
  ggplot(mapping = aes(x = goodFood) ) +
  geom_density() +
  theme_bw() +
  labs(title = "Density of good food ratings",
       x = "Rating on 1-10 scale",
       y = “Density")
```

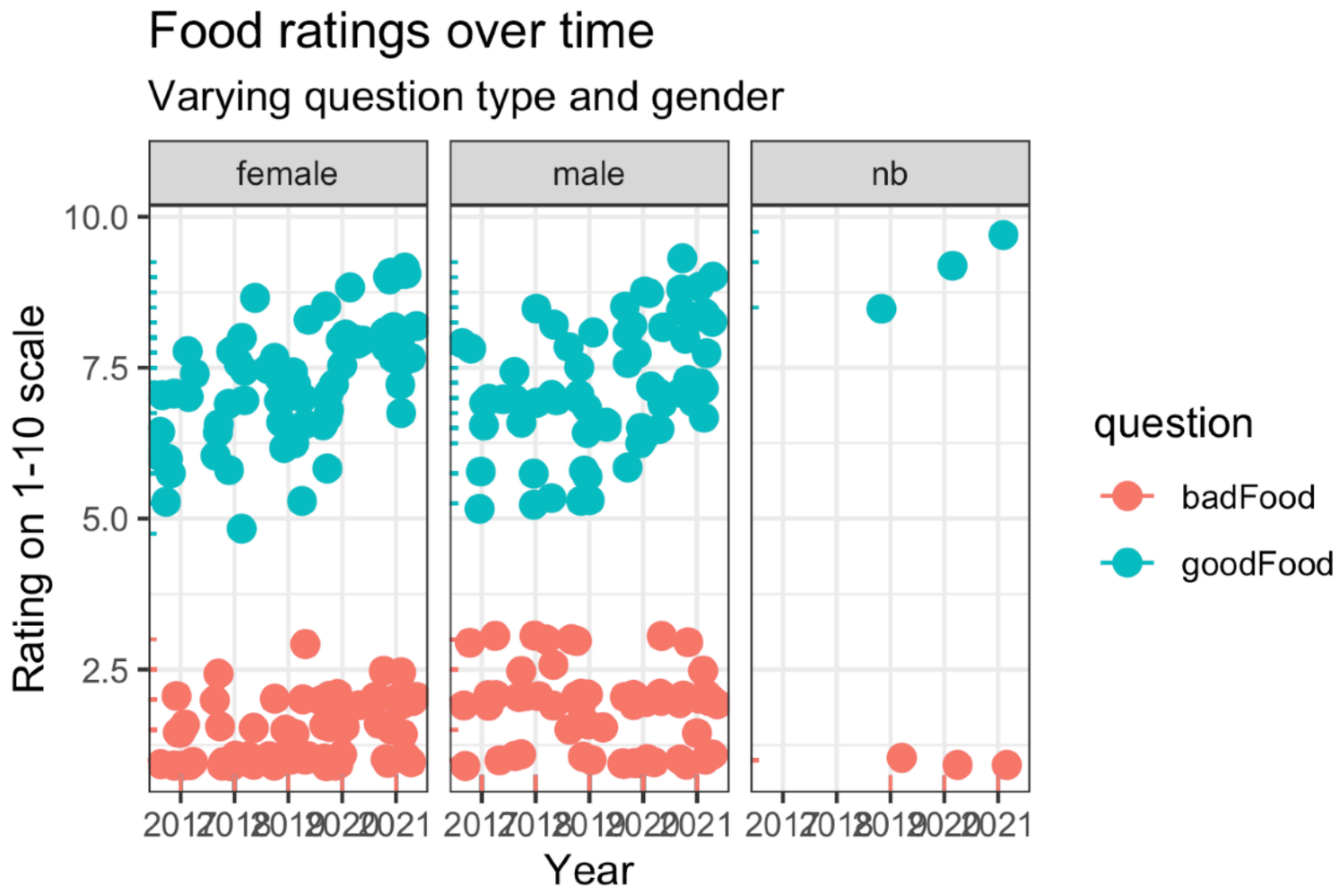


There are also a lot of two-variable geoms

Two variables

We've already seen some of these:

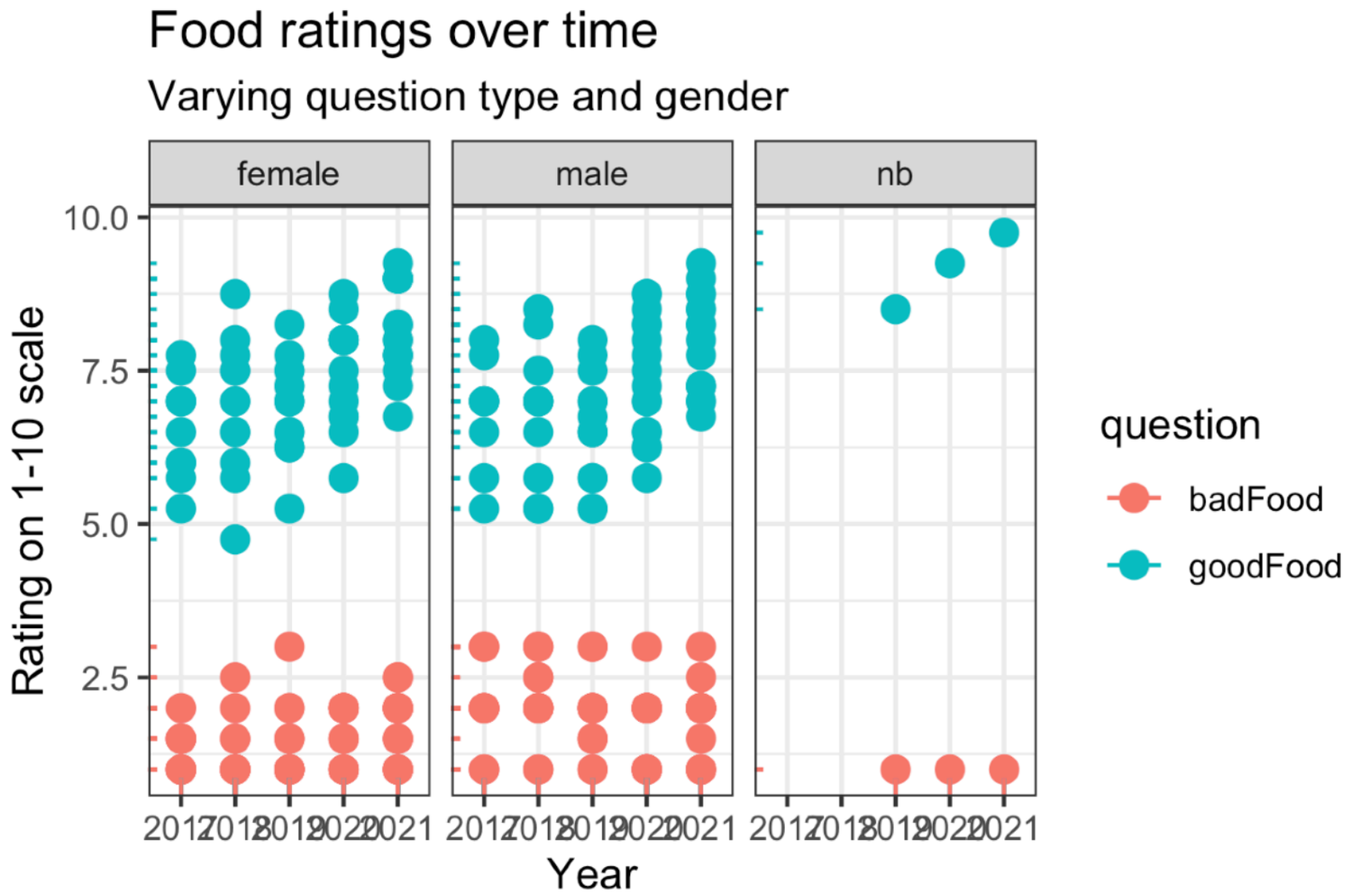
```
dl %>%
  ggplot(
    mapping = aes(
      x = year,
      y = rating,
      colour = question
    )
  ) +
geom_jitter(size=3) +
geom_rug() +
facet_wrap(~gender) +
theme_bw() +
labs(
  title = "Food ratings over time",
  subtitle = "Varying question type and gender",
  x = "Year",
  y = "Rating on 1-10 scale"
)
```



Two variables

We've already seen some of these:

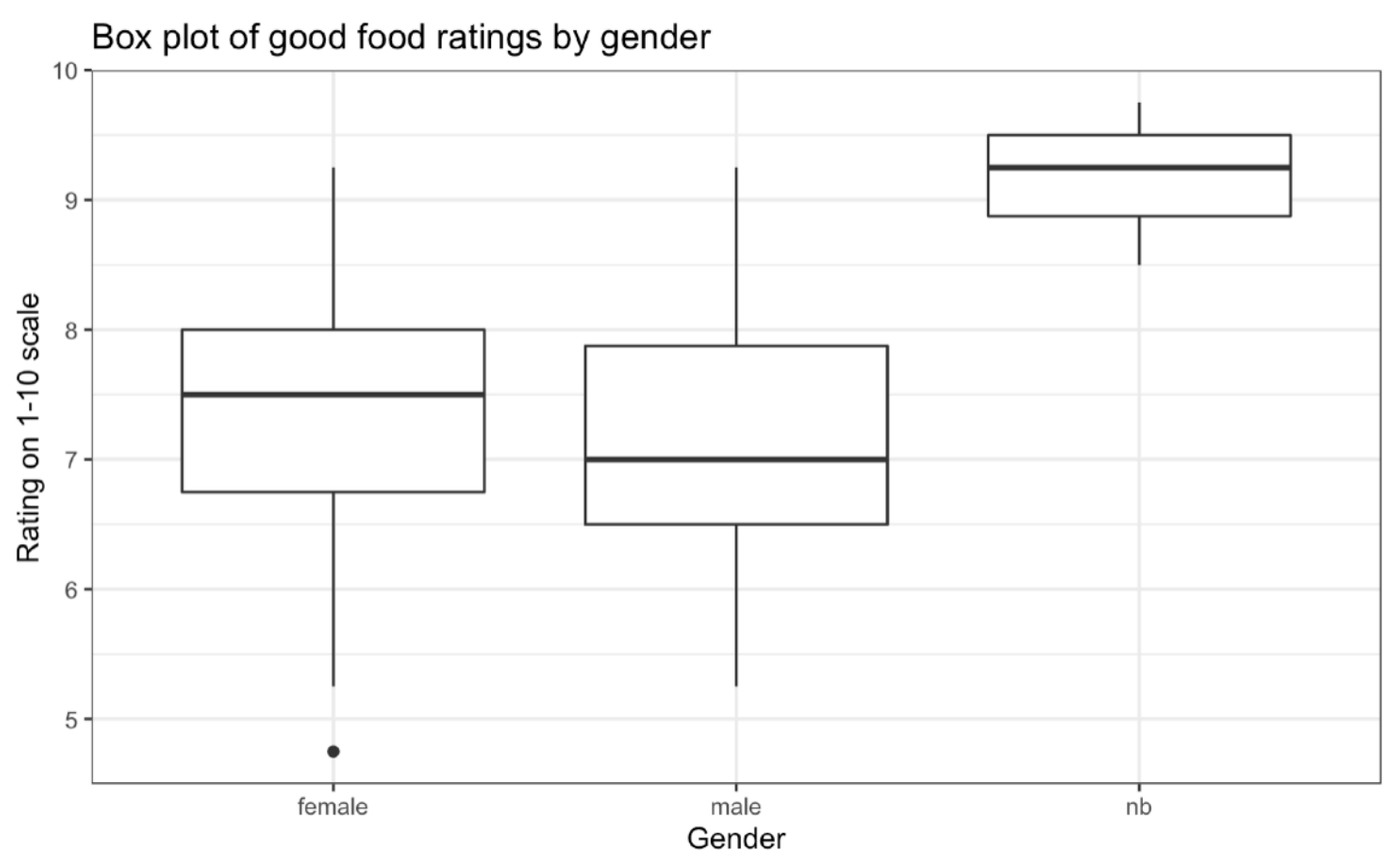
```
dl %>%
  ggplot(
    mapping = aes(
      x = year,
      y = rating,
      colour = question
    )
  ) +
geom_point(size=3) +
geom_rug() +
facet_wrap(~gender) +
theme_bw() +
labs(
  title = "Food ratings over time",
  subtitle = "Varying question type and gender",
  x = "Year",
  y = "Rating on 1-10 scale"
)
```



Two variables, one discrete

Box plots: summarise medians, IQR, etc

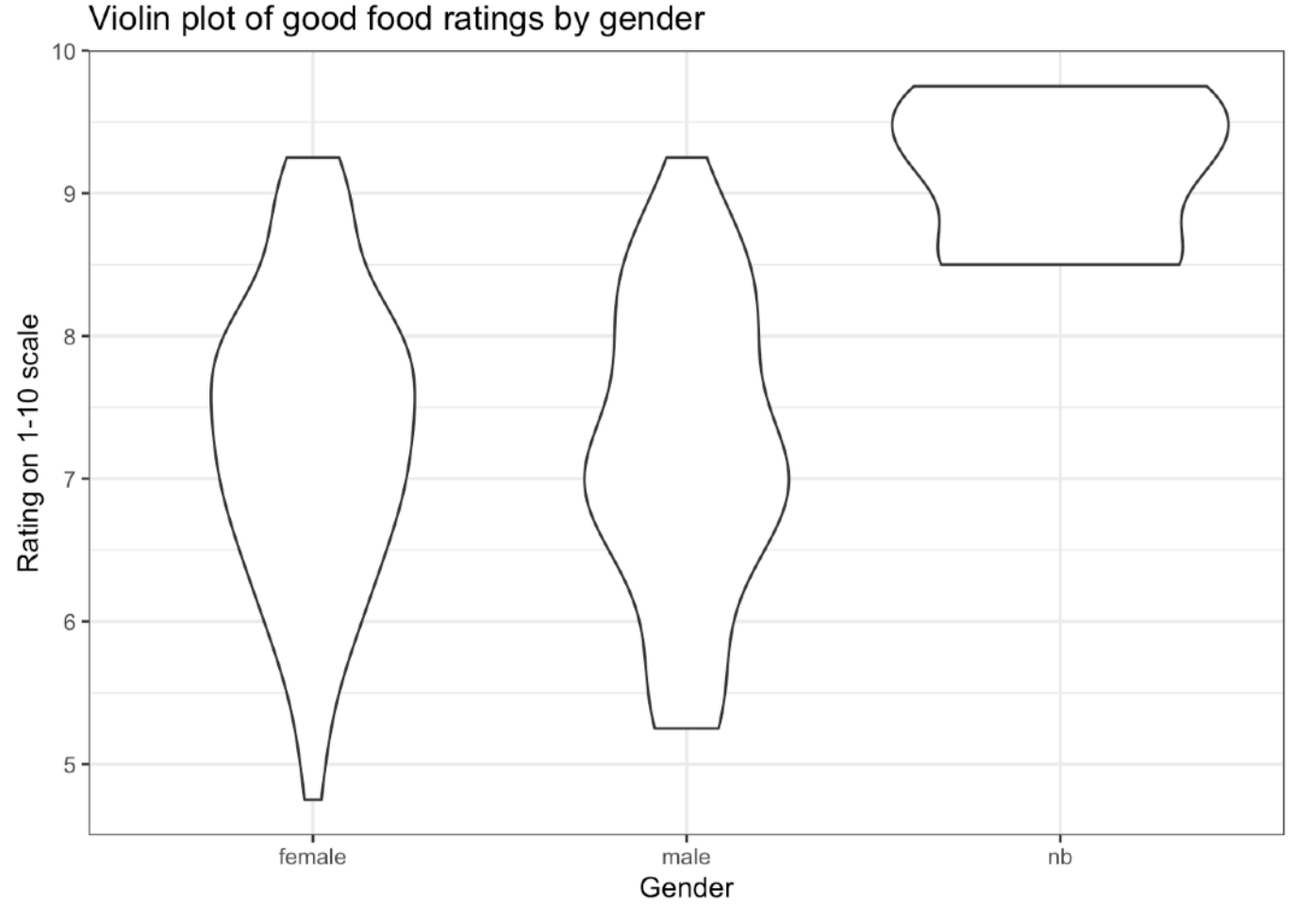
```
dnew %>%
  ggplot(mapping = aes(x = gender, y=goodFood)) +
  geom_boxplot() +
  theme_bw() +
  labs(title = "Boxplot of good food ratings",
       x = "Rating on 1-10 scale",
       y = “Gender")
```



Two variables, one discrete

Violin plots: give sense of distributional shape

```
dnew %>%
  ggplot(mapping = aes(x = gender, y=goodFood)) +
  geom_violin() +
  theme_bw() +
  labs(title = “Violin plot of good food ratings",
       x = "Rating on 1-10 scale",
       y = “Gender")
```



See the `w4day1exercises.Rmd` file for the exercises!