Visualisation:
Adding coolness to graphs

Research Methods for Human Inquiry

Andrew Perfors

So far we've seen the basics of everything you need to make figures

But today we're going to add all of the coolness

- Multiple geoms in a single plot
- Colour palettes

Adding coolness

Let's start by loading up our data. Click on the w4day2 project icon in the w4day2 zipped file you downloaded from Canvas. Then open `w4day2analysis.Rmd`:

It loads up everything we need and creates the tibbles in the right form as we did in Day 1

Multiple geoms

Remember how making figures is like an artist doing a painting in different layers? Each geom is a different layer. So let's build something cool up using `dl` (from last time)

A tibble: 270 × 6

name	gender	species	year	question	rating
<chr>	<chr>	<chr>	<fctr>	<chr>	<dbl>
foxy	female	fox	2021	goodFood	7.25
foxy	female	fox	2021	badFood	1.00
bunny	female	bunny	2021	goodFood	9.00
bunny	female	bunny	2021	badFood	1.00
doggie	male	dog	2021	goodFood	8.75
doggie	male	dog	2021	badFood	1.00
flopsy	nb	bunny	2021	goodFood	9.75
flopsy	nb	bunny	2021	badFood	1.00
fluffy	female	cat	2021	goodFood	8.00
fluffy	female	cat	2021	badFood	1.50

1–10 of 270 rows Previous 1 2 3 4 5 6 ... 27 Next

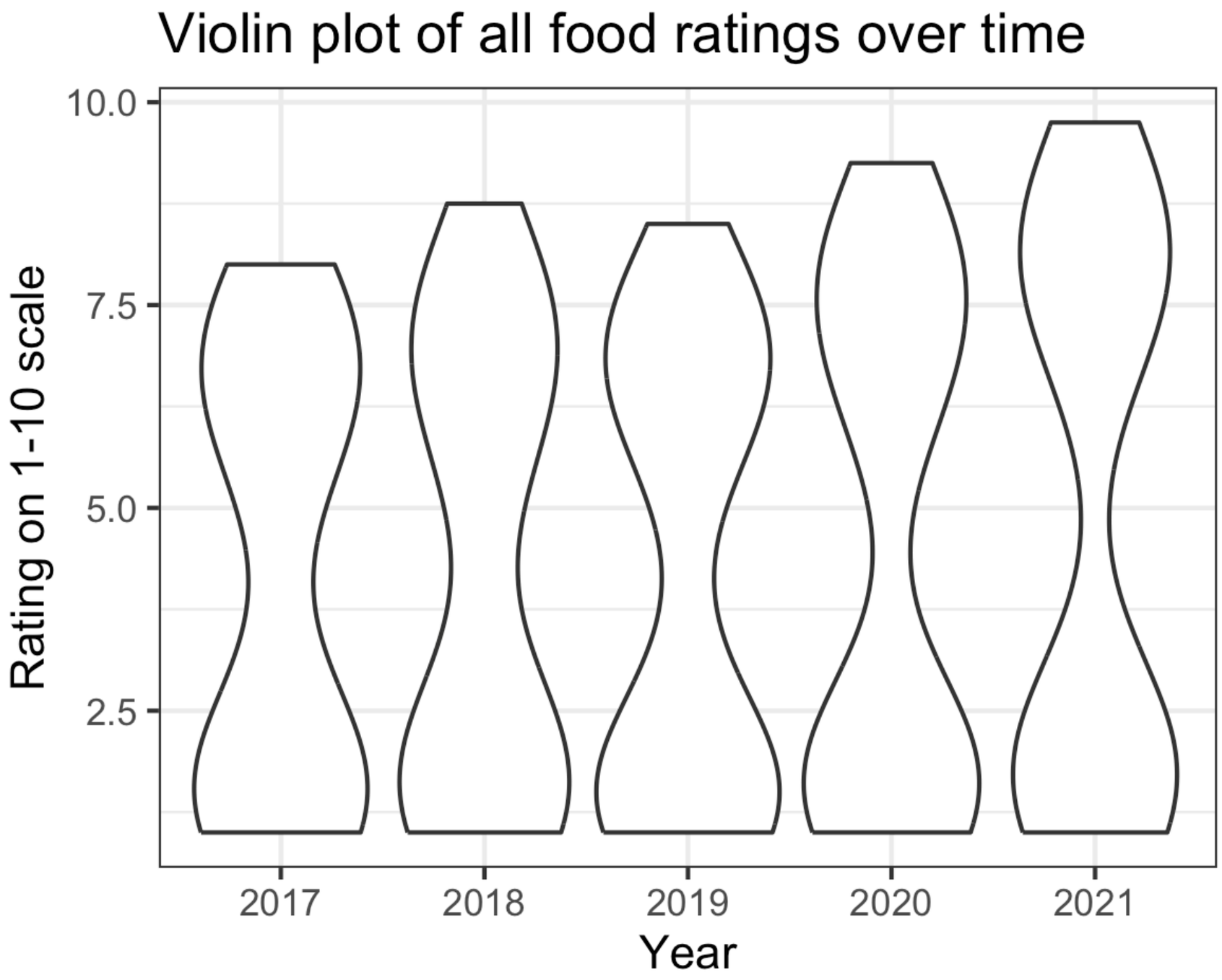
Multiple geoms

First let's see how ratings change over time:

```
dl %>%
  ggplot(mapping = aes(x = year, y = rating)) +
  geom_violin() +
  theme_bw() +
  labs(title = "Violin plot of all food ratings over time",
  y = "Rating on 1-10 scale",
  x = "Year")
```

Multiple geoms

First let's see how ratings change over time:



A bit odd — probably because ratings includes both `goodFood` and `badFood`!

Multiple geoms

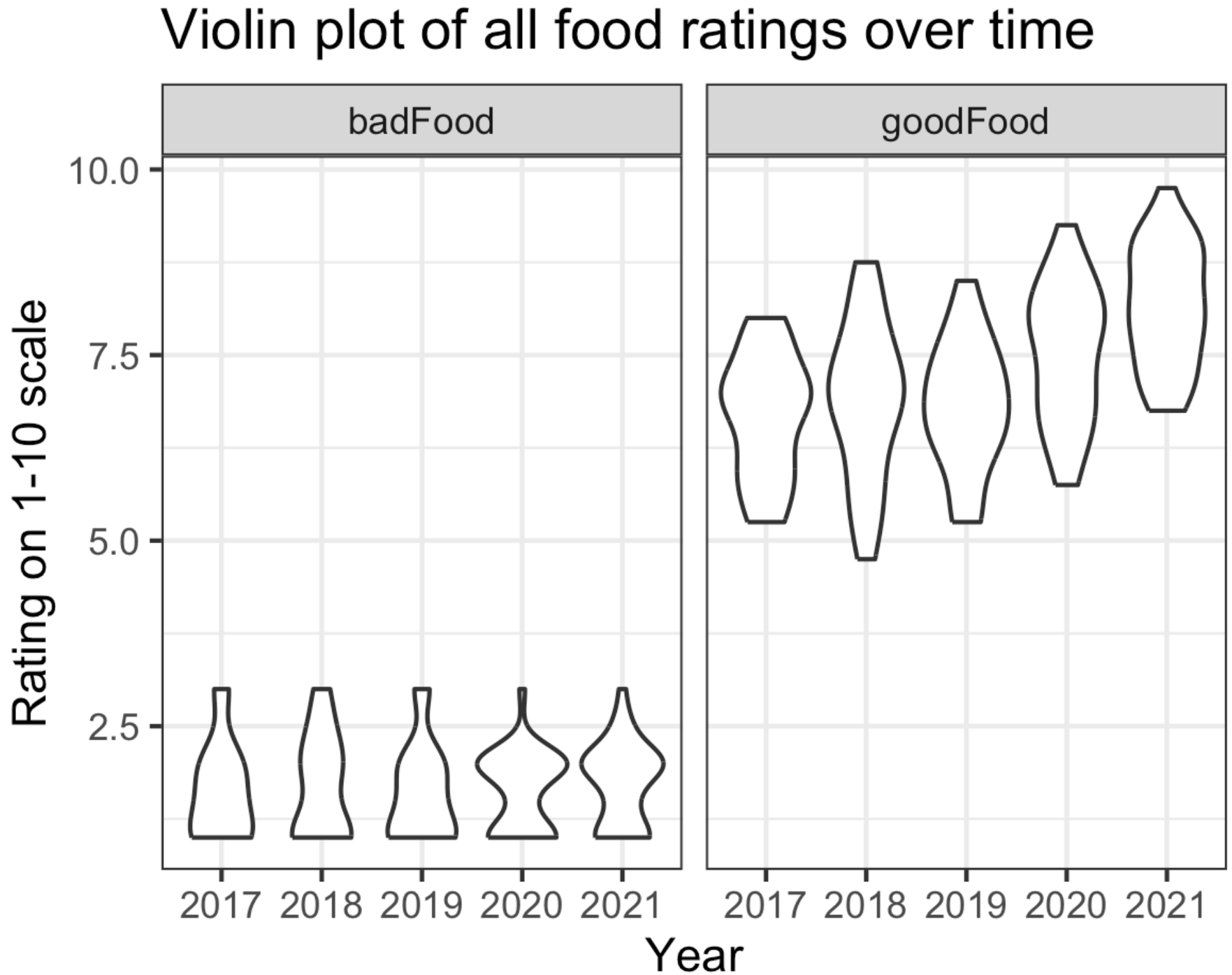
Split goodFood and badFood into facets/panels

```
dl %>%
  ggplot(mapping = aes(x = year, y = rating)) +
  geom_violin() +
  facet_wrap(~question) +
  theme_bw() +
  labs(title = "Violin plot of all food ratings over time",
    y = "Rating on 1-10 scale",
    x = "Year")
```

Multiple geoms

Split goodFood and badFood into facets/panels

This is far more sensible, and seems to suggest that ratings for good foods are going up but bad foods aren't! Hmm… why would that be?



For a start, this is more evidence that people know how to use the scale correctly — answers for the bad foods are not only different than for the good foods, they aren't increasing over time.

I knew it!!!
We're all starving to death!!!
OMG OMG OMG


Maybe this has nothing to do with hunger — what was the exact question asked again?

"On a scale of 1-10, how much would you want to eat a XX right now?"


Could be hunger! I'm hungry.

Let's look more closely..

Let's see if these distributions are hiding detail by adding on the specific datapoints, and add colours so we see them better

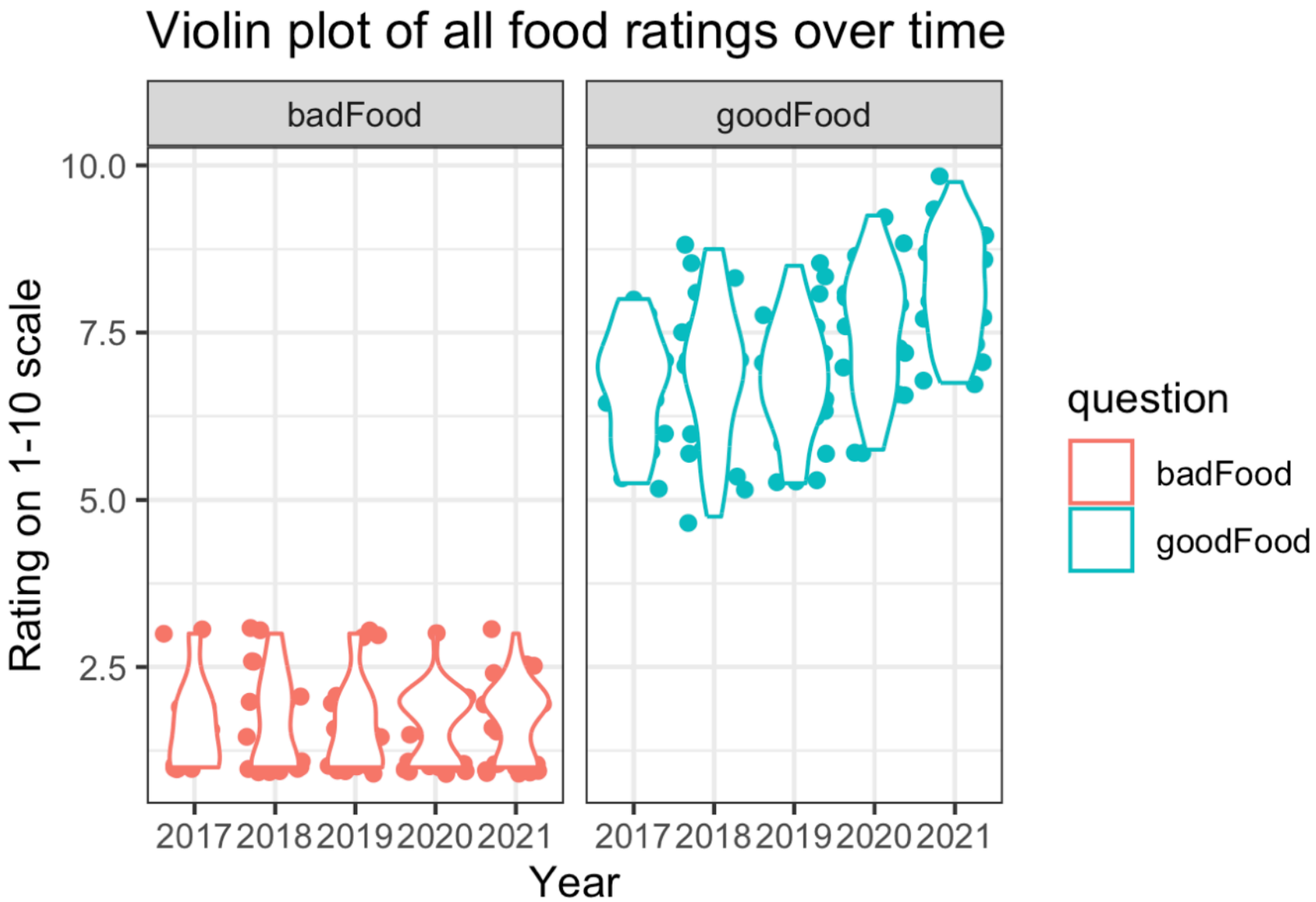
```
dl %>%
  ggplot(mapping = aes(x = year, y = rating, colour=question)) +
  geom_jitter() +
  geom_violin() +
  facet_wrap(~question) +
  theme_bw() +
  labs(title = "Violin plot of all food ratings over time",
    y = "Rating on 1-10 scale",
    x = "Year")
```

Let’s look more closely..

Let’s see if these distributions are hiding detail by adding on the specific datapoints, and add colours so we see them better

Oops! The datapoints are partly hidden by the violin (since we added them first). Let’s bring them to the front but make them semi-transparent

Also, the legend is redundant and ugly, so let’s get rid of it



Let's look more closely..

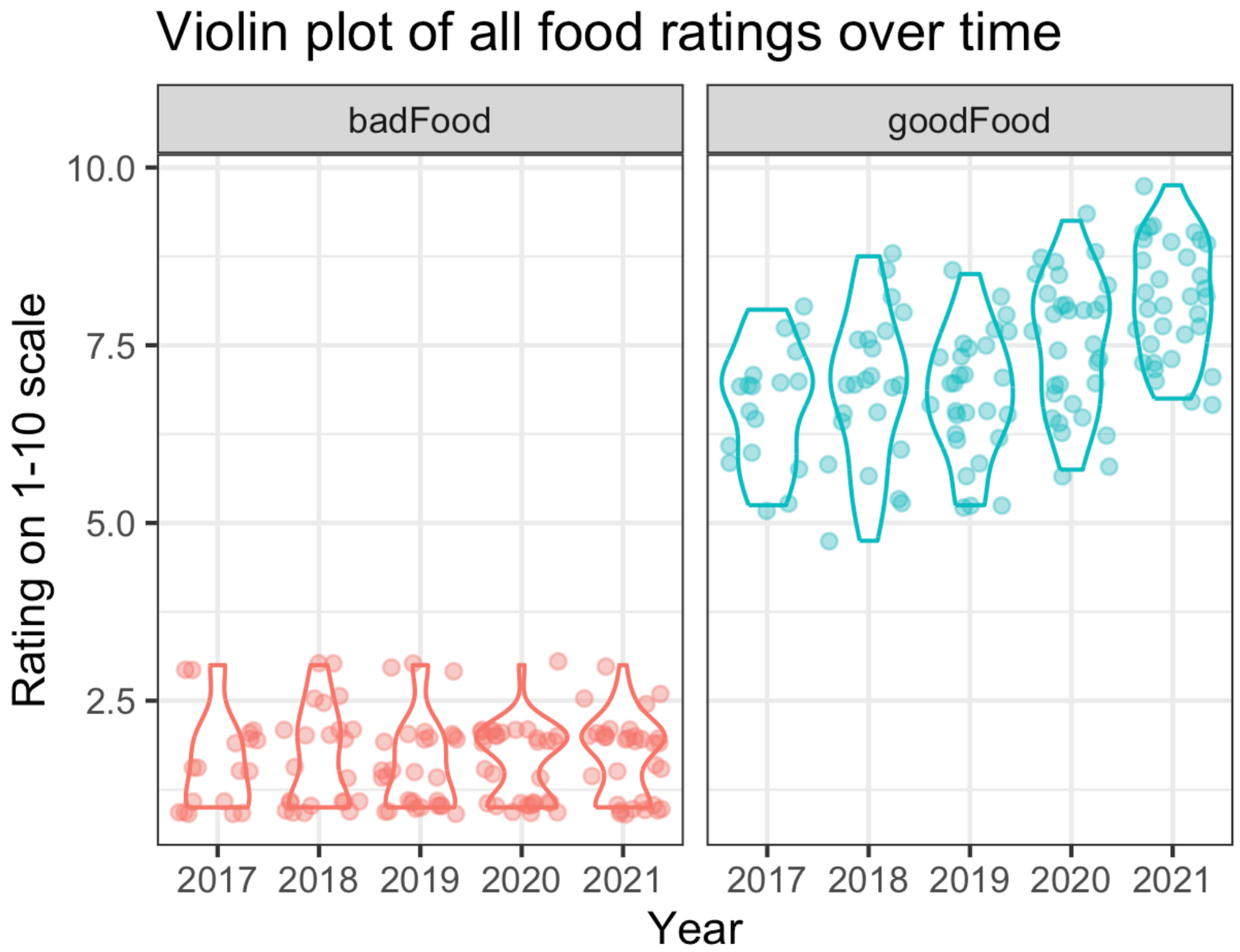
Need to remove the legend out of both geoms, put dots in front, and make them semi-transparent

```
dl %>%
  ggplot(mapping = aes(x = year, y = rating, colour=question)) +
  geom_violin(show.legend=FALSE) +
  geom_jitter(show.legend=FALSE,alpha=0.4) +
  facet_wrap(~question) +
  theme_bw() +
  labs(title = "Violin plot of all food ratings over time",
    y = "Rating on 1-10 scale",
    x = "Year")
```

Let’s look more closely..

Need to remove the legend out of both geoms, put dots in front, and make them semi-transparent

Looks nice, and we can now see clearly that goodFood ratings are going up with time but badFood are not



Hey, no offence, but these colours kind of suck. Can you do anything about that?

Making better colours

There are two kinds of ways to add colour to a graph: as a `fill` (for filling in shapes) or as a `colour` (for lines)

To demonstrate, let's first reorganise our original dataset so we have a single variable called `question`:

```
dl2 <- d %>%
  pivot_longer(-c(name,gender,species,year),names_to="question",values_to="rating")
```

A tibble: 810 × 6

name <chr>	gender <chr>	species <chr>	year <fctr>	question <chr>	rating <dbl>
foxy	female	fox	2021	carrot	7
foxy	female	fox	2021	bean	7
foxy	female	fox	2021	cake	8
foxy	female	fox	2021	meat	7
foxy	female	fox	2021	mud	1
foxy	female	fox	2021	dirt	1
bunny	female	bunny	2021	carrot	10

Making better colours

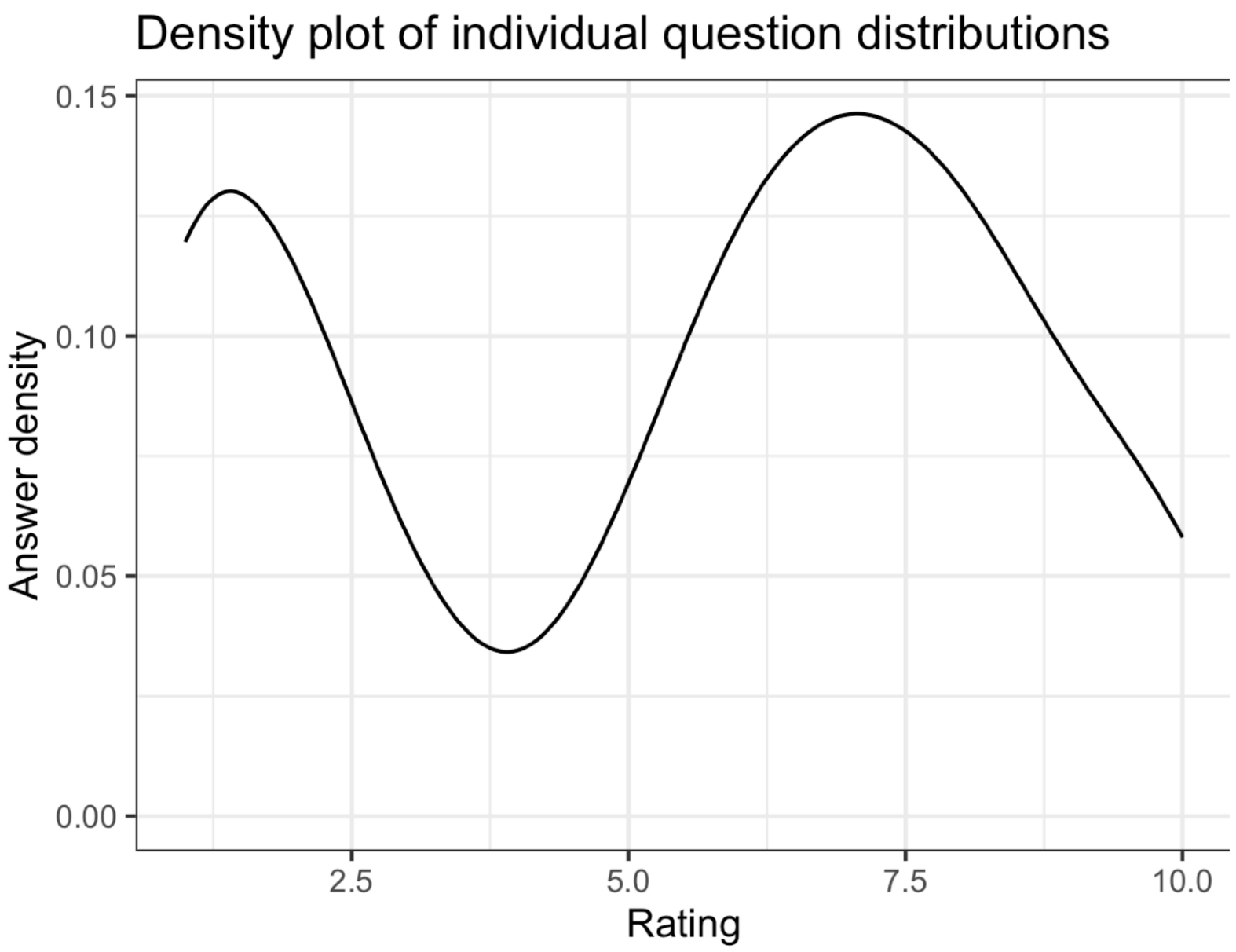
Now let's make a density plot to see the answers

```
dl2 %>%
  ggplot(mapping = aes(x=rating)) +
  geom_density() +
  theme_bw() +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

Making better colours

Now let's make a density plot to see the answers

This does not show individual questions! It shows all of the ratings for all of the questions combined together



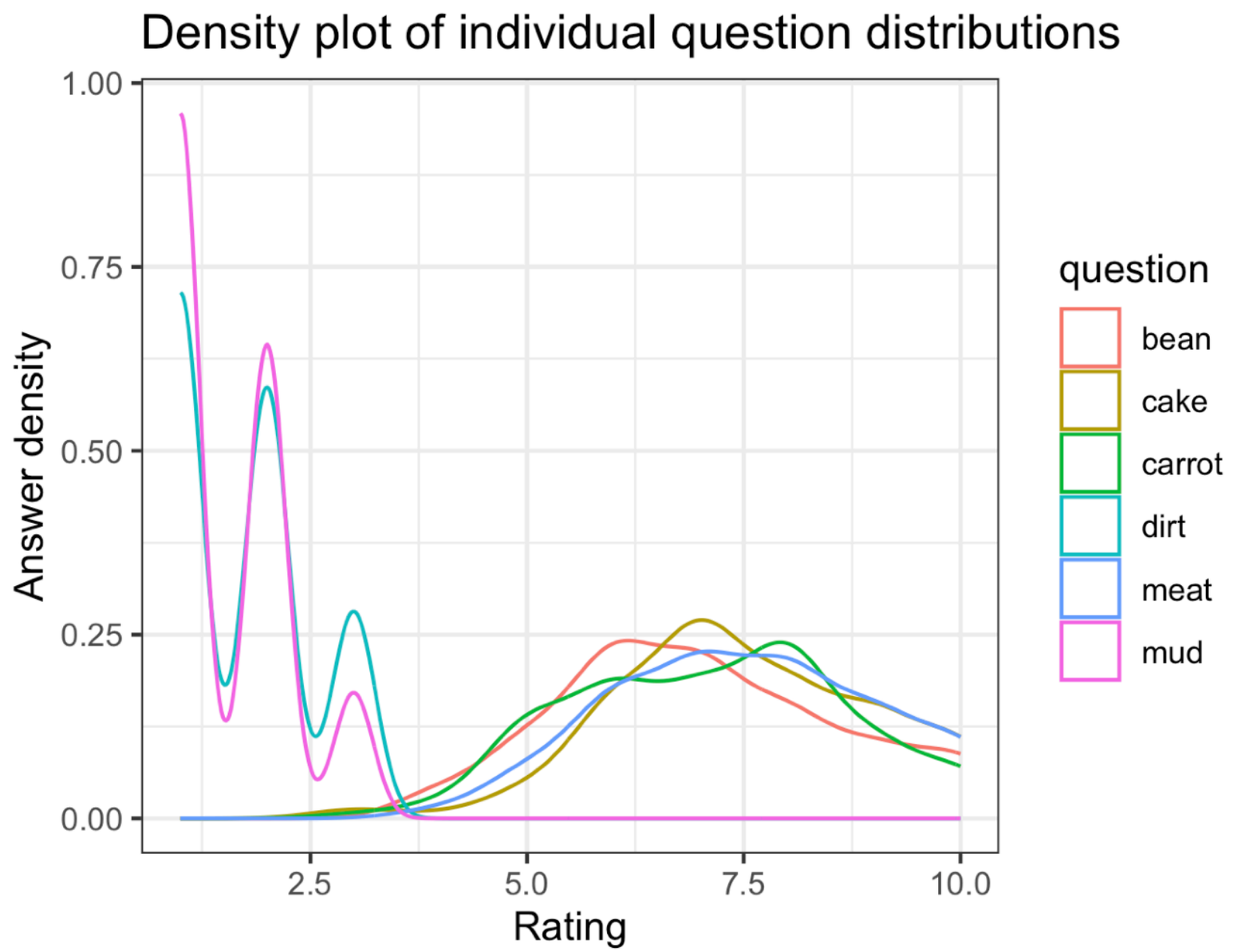
Making better colours

Now let's make a density plot to see the answers for each question.

```
dl2 %>%
  ggplot(mapping = aes(x=rating,colour=question)) +
  geom_density() +
  theme_bw() +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

Making better colours

Now let's make a density plot to see the answers for each question.



Note that this makes the lines for each question a different colour.

Making better colours

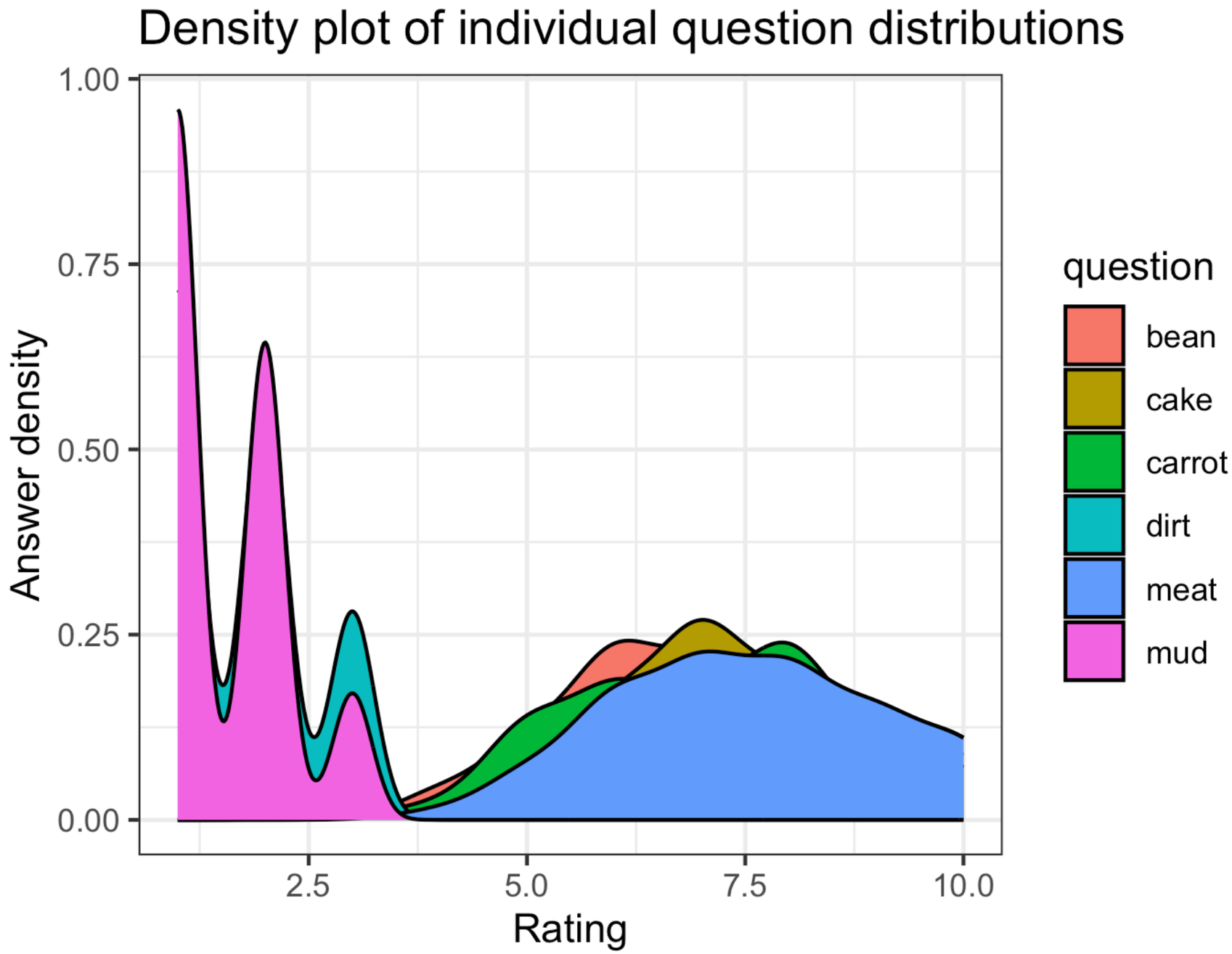
Let's fill the plot with colours and make the lines black

```
dl2 %>%
  ggplot(mapping = aes(x=rating,fill=question)) +
  geom_density(colour="black") +
  theme_bw() +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

Note that this makes the lines for each question black.

Making better colours

Let's fill the plot with colours and make the lines black



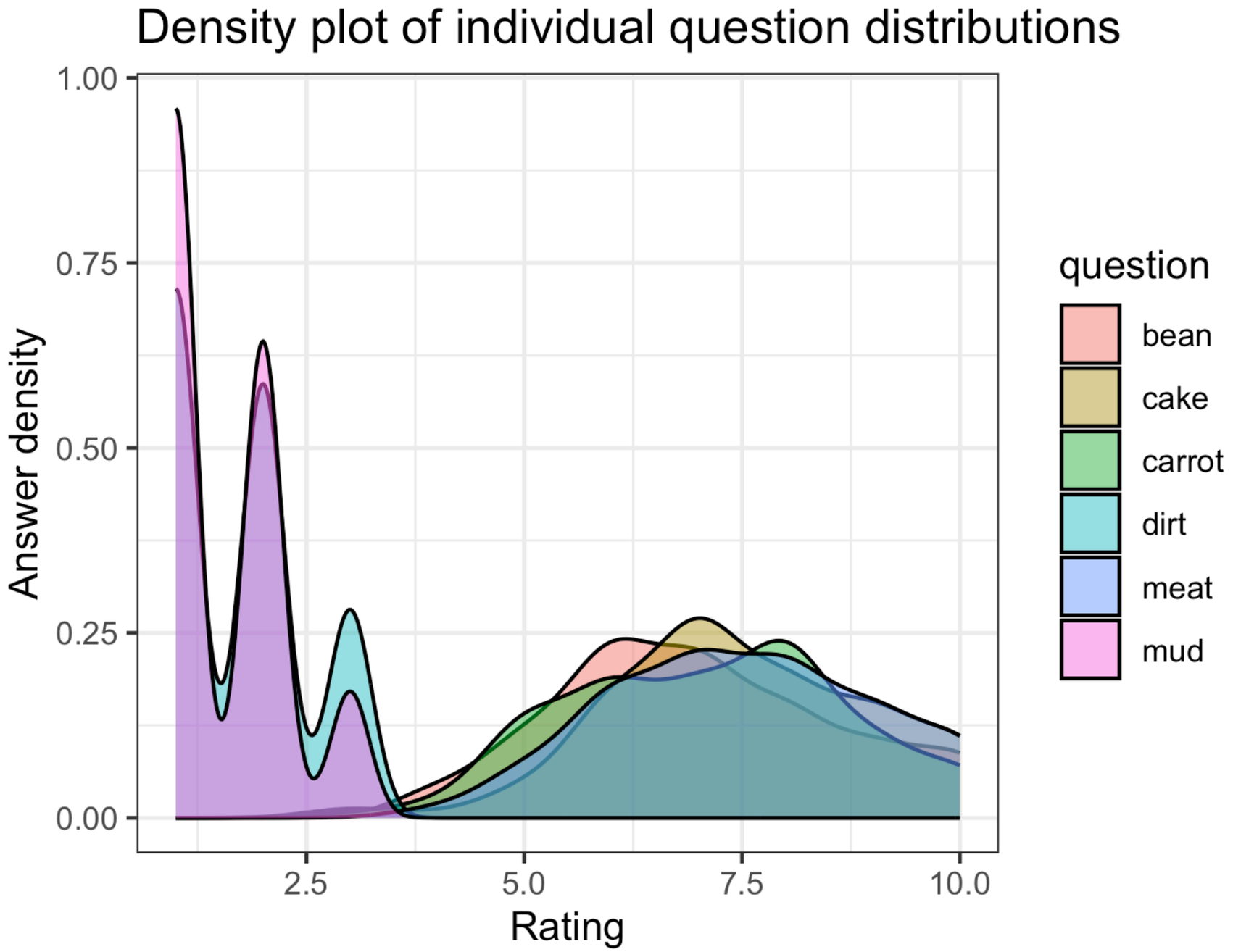
Making better colours

Use alpha to make the fill slightly transparent

```
dl2 %>%
  ggplot(mapping = aes(x=rating,fill=question)) +
  geom_density(colour="black",alpha=0.5) +
  theme_bw() +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

Making better colours

Use alpha to make the fill slightly transparent



You can change the default colours with palettes

There are lots of different palettes available. One good one is called ColorBrewer. Install the package corresponding to it

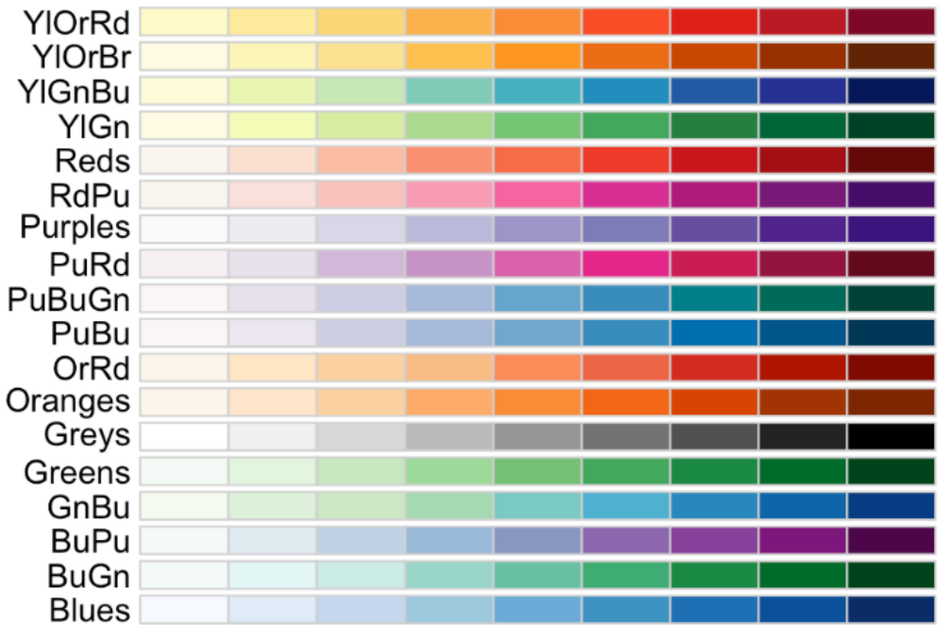
```
install.packages(“RColorBrewer”)
library(RColorBrewer)
```

RColorBrewer has a range of palettes designed to look nice. They are mostly defined for discrete data, but can be extended to continuous data.

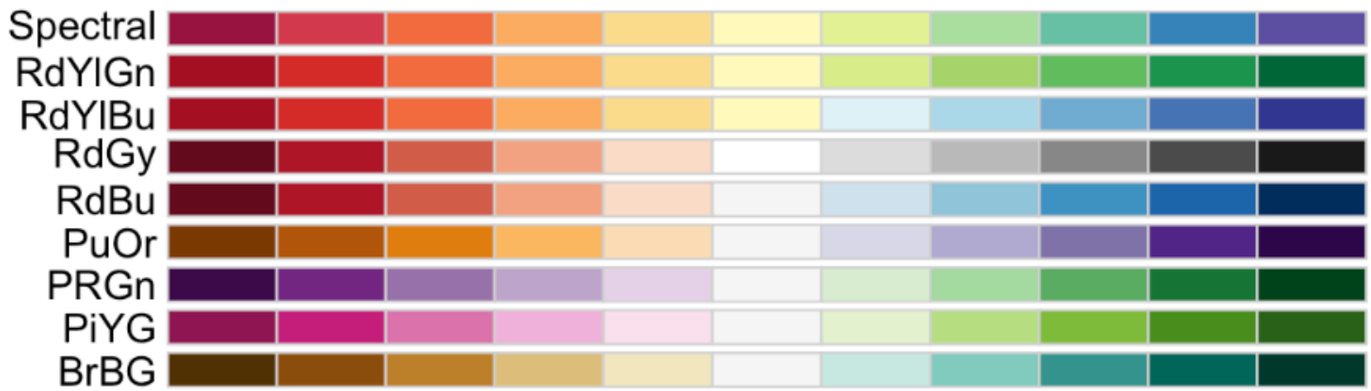
https://colorbrewer2.org/

RColorBrewer

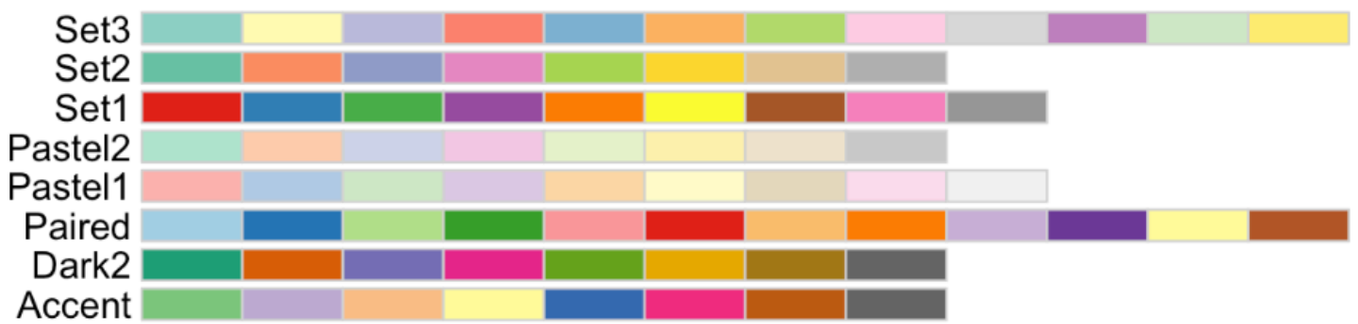
Sequential: ordered data that progresses from low to high



Qualitative: good for categorical data with different classes



Diverging: All parts of the range, similar contrast throughout



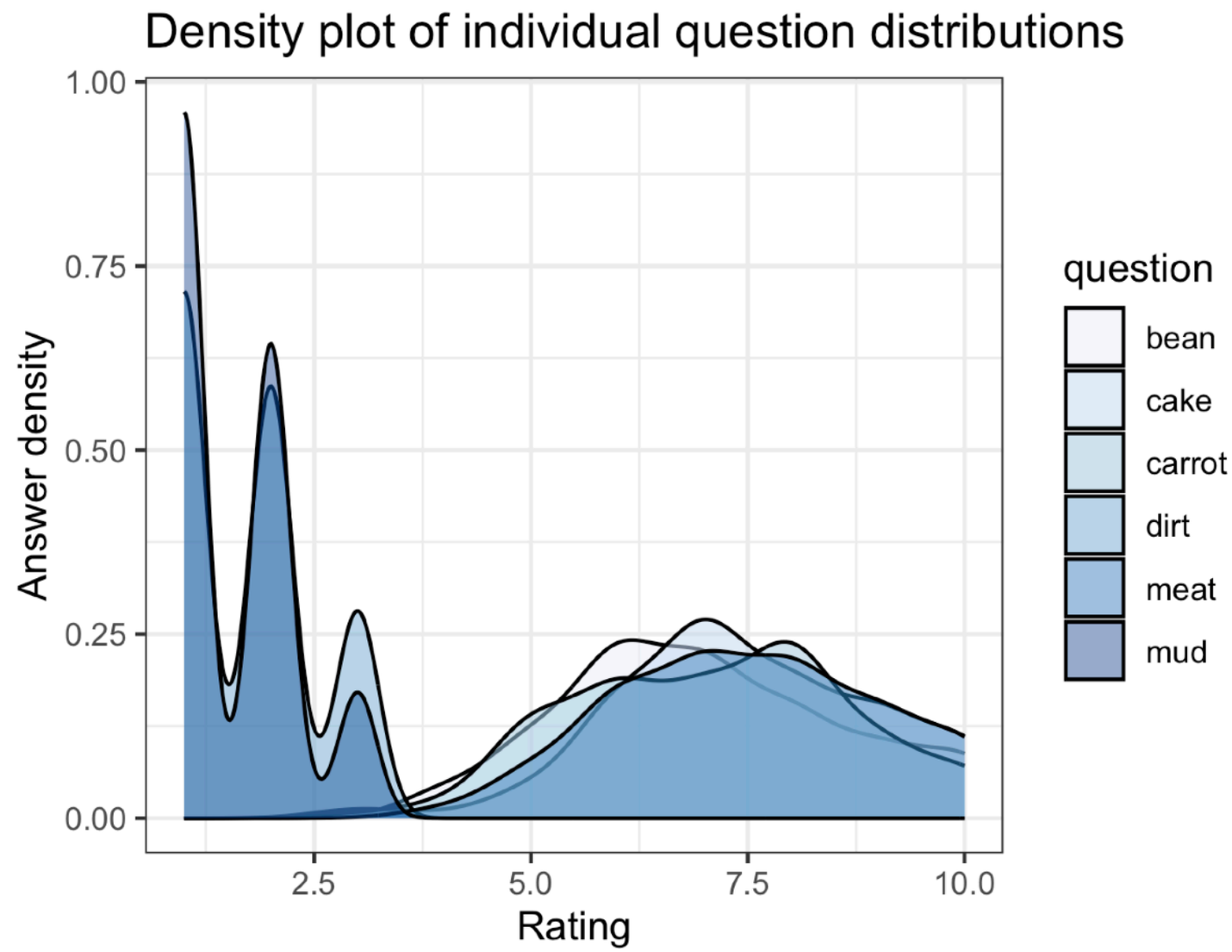
You can change the default colours with palettes

If it's a fill, use the `scale_fill_brewer()` function with the palette name as an argument.

```
dl2 %>%
  ggplot(mapping = aes(x=rating,fill=question)) +
  geom_density(colour="black",alpha=0.5) +
  theme_bw() +
  scale_fill_brewer(palette="Blues") +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

You can change the default colours with palettes

If it's a fill, use the `scale_fill_brewer()` function with the palette name as an argument.



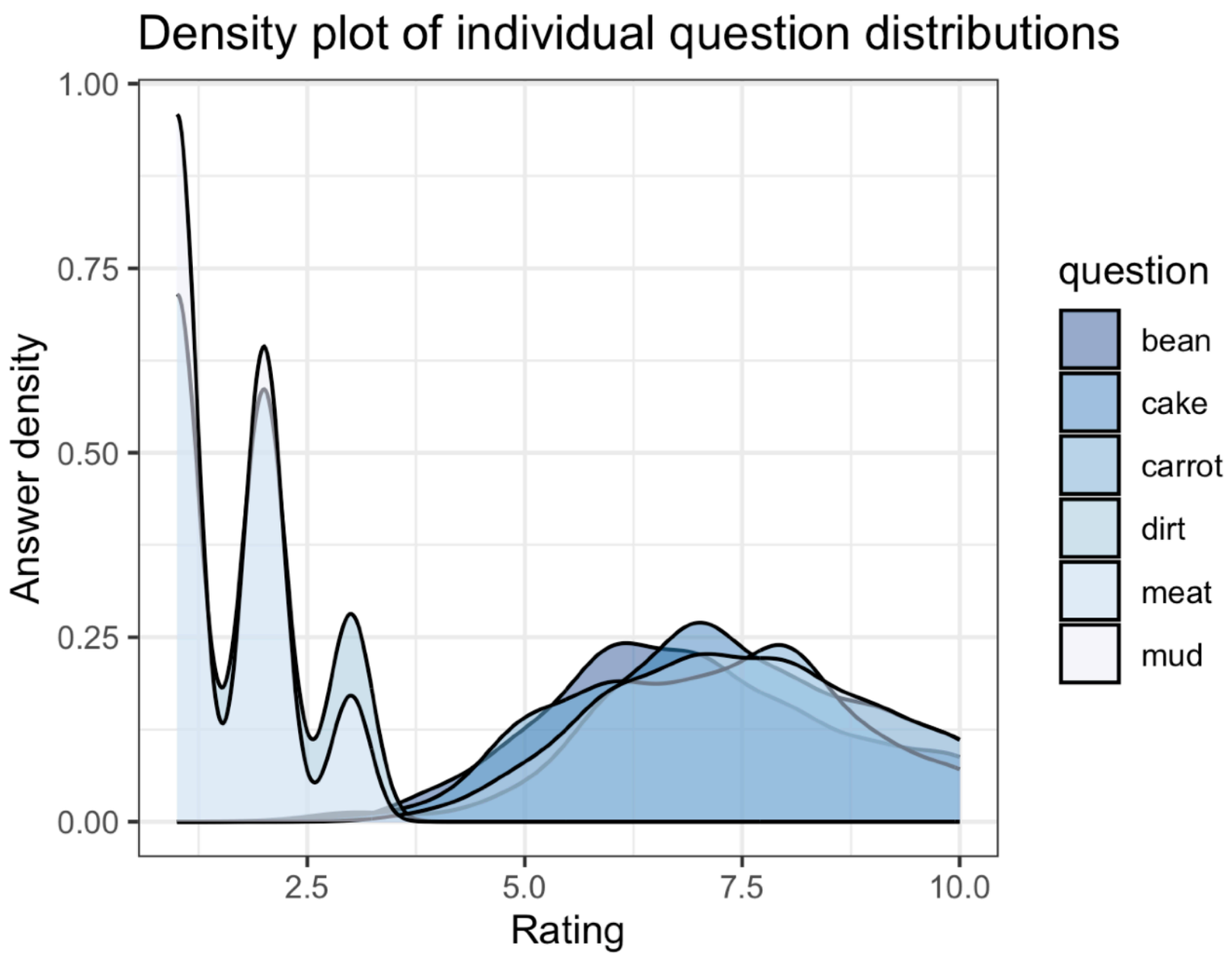
You can change the default colours with palettes

For sequential palettes, you can change the direction of how they map onto variables

```
dl2 %>%
  ggplot(mapping = aes(x=rating,fill=question)) +
  geom_density(colour="black",alpha=0.5) +
  theme_bw() +
  scale_fill_brewer(palette="Blues", direction=-1) +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

You can change the default colours with palettes

For sequential palettes, you can change the direction of how they map onto variables



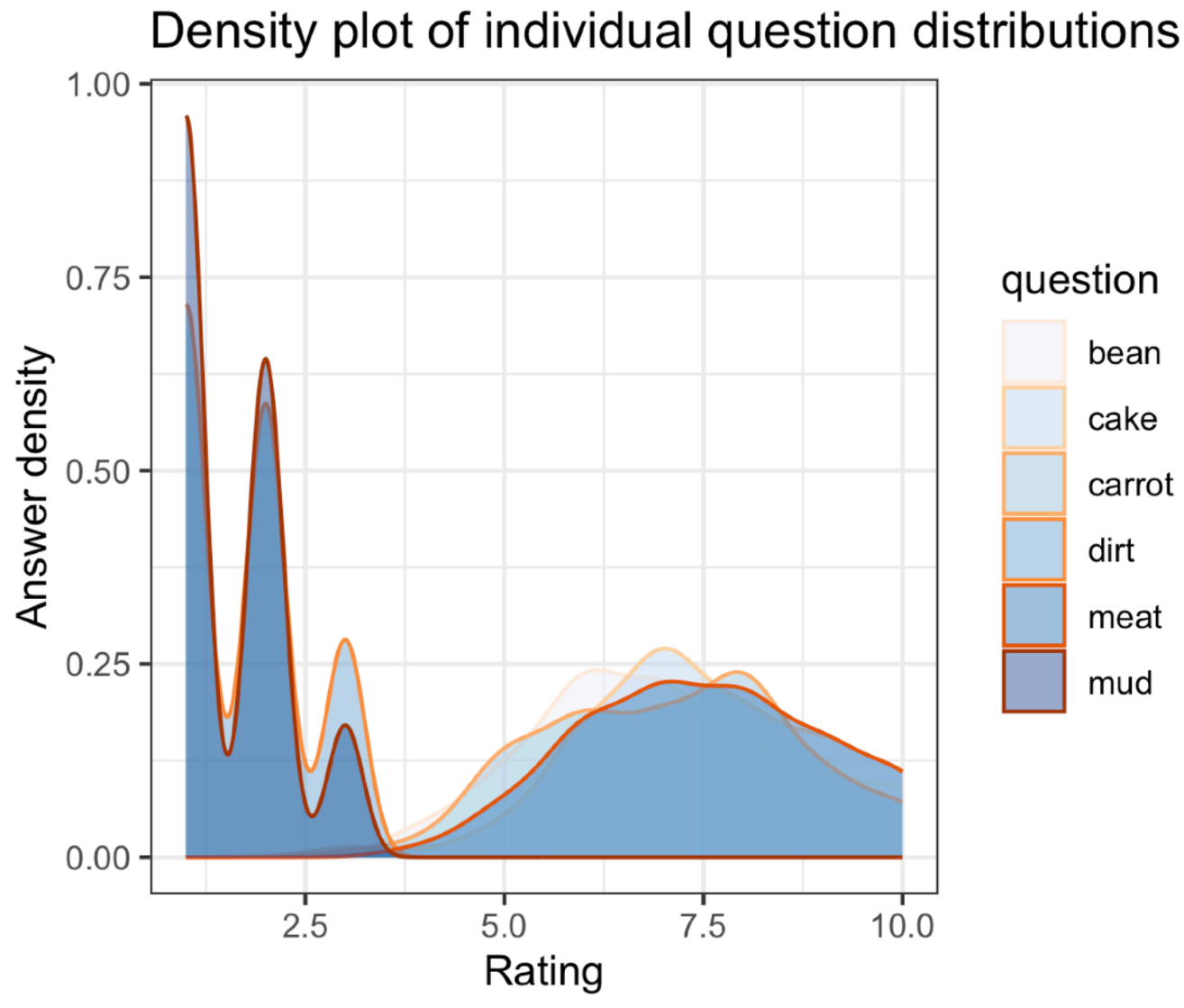
You can change the default colours with palettes

If it's a line, use the `scale_color_brewer()` function with the palette name as an argument.

```
dl2 %>%
  ggplot(mapping = aes(x=rating,fill=question,colour=question)) +
  geom_density(alpha=0.5) +
  theme_bw() +
  scale_fill_brewer(palette="Blues") +
  scale_color_brewer(palette="Oranges") +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

You can change the default colours with palettes

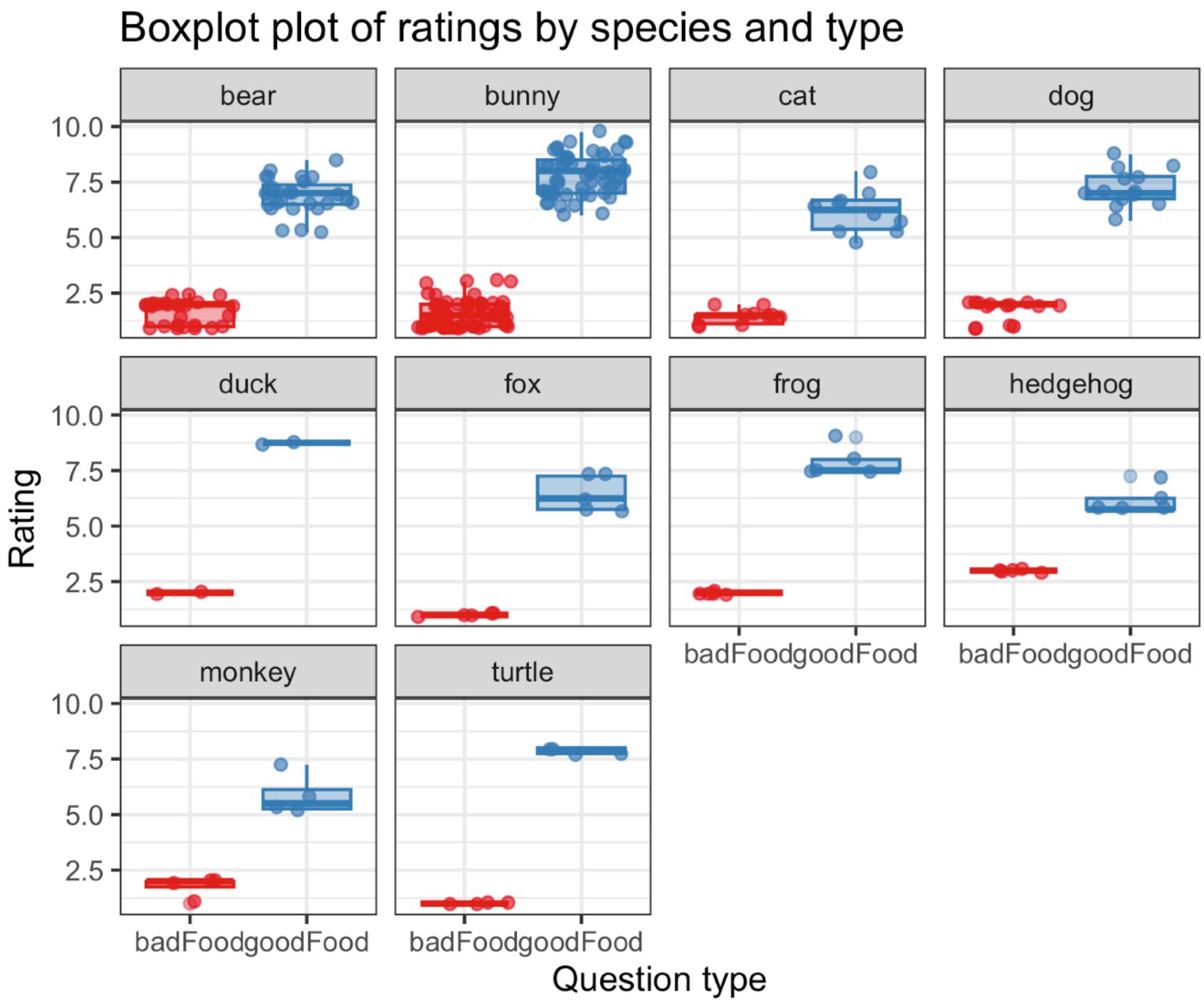
If it's a line, use the `scale_color_brewer()` function with the palette name as an argument.



Combine coolness now!!

```
dl %>%
  ggplot(mapping = aes(x = question, y = rating,
      fill = question, colour = question)) +
  geom_boxplot(alpha=0.4,show.legend=FALSE) +
  geom_jitter(alpha=0.7,show.legend=FALSE) +
  scale_fill_brewer(palette="Set1") +
  scale_color_brewer(palette="Set1") +
  facet_wrap(~species) +
  theme_bw() +
  labs(title = "Boxplot plot of ratings by species and type",
    y = "Rating",
    x = "Question type"
  )
```

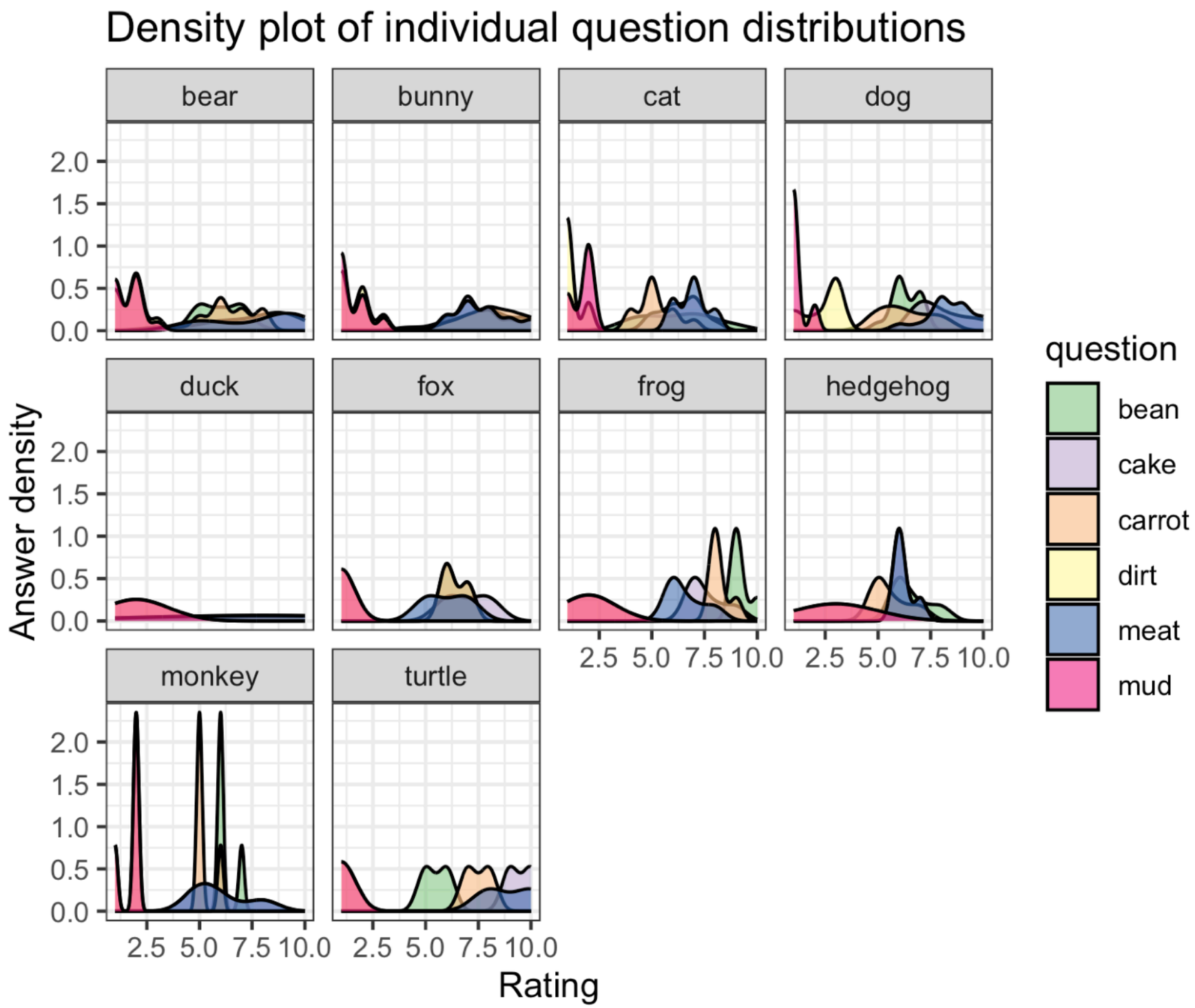
Combine coolness now!!



Combine coolness now!!

```
dl2 %>%
  ggplot(mapping = aes(x = rating,fill = question)) +
  geom_density(alpha=0.6,colour="black") +
  scale_fill_brewer(palette="Accent") +
facet_wrap(~species) +
theme_bw() +
labs(title = "Density plot of individual question distributions",
  y = "Answer density",
  x = "Rating")
```

Combine coolness now!!

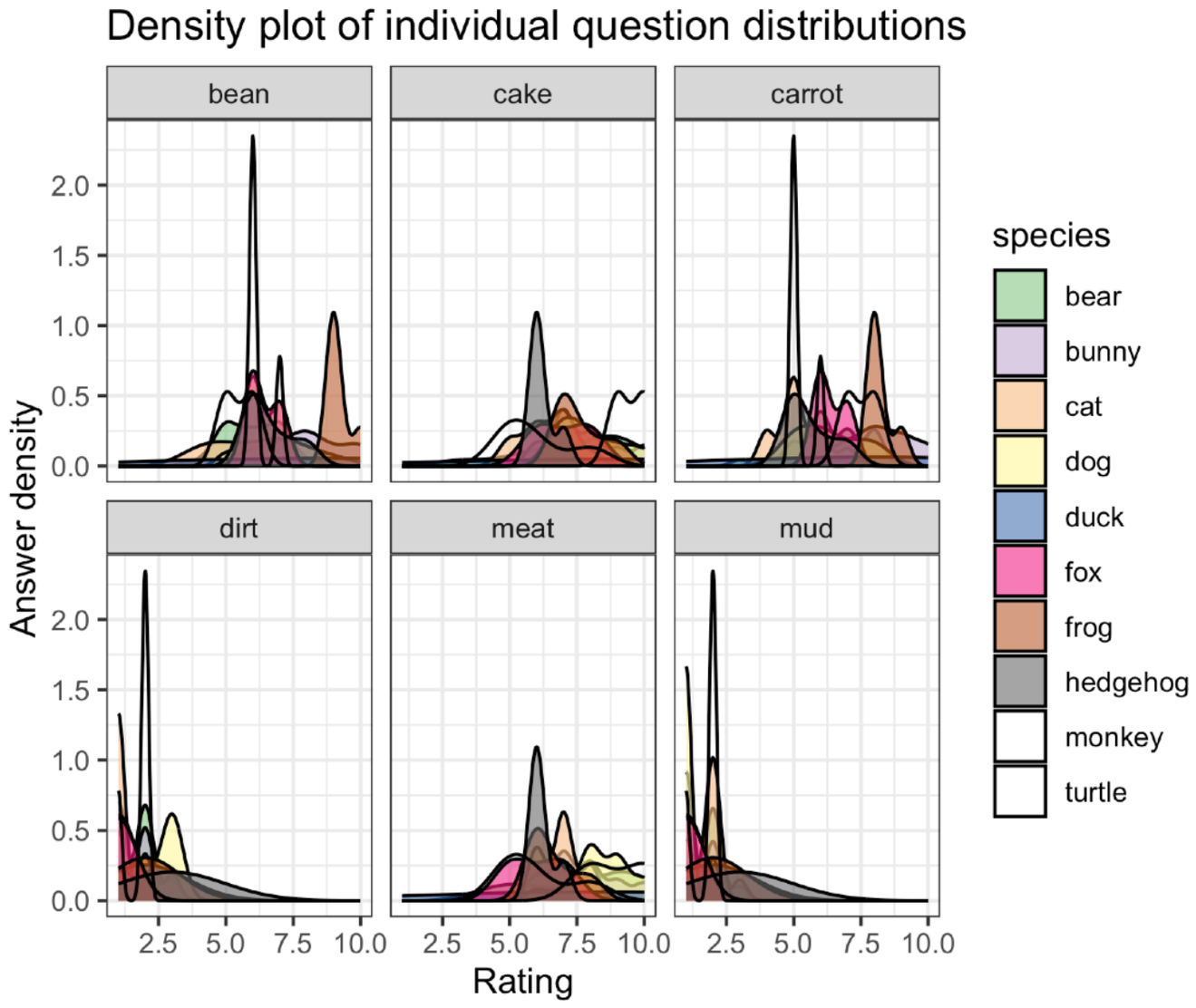


Combine coolness now!!

```
dl2 %>%
  ggplot(mapping = aes(x = rating,fill = species)) +
  geom_density(alpha=0.6,colour="black") +
  scale_fill_brewer(palette="Accent") +
  facet_wrap(~question) +
  theme_bw() +
  labs(title = "Density plot of individual question distributions",
    y = "Answer density",
    x = "Rating")
```

Combine coolness now!!

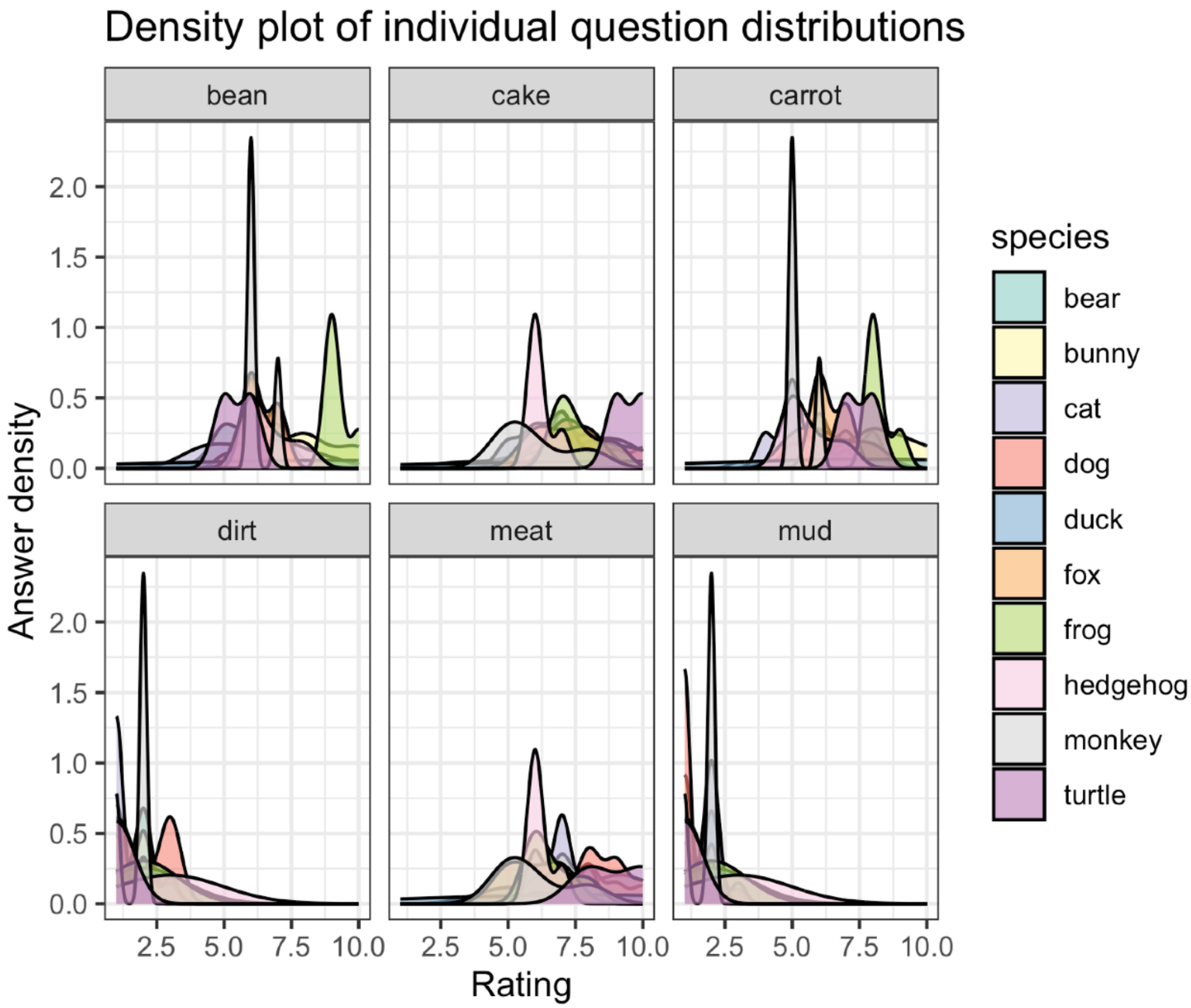
Palettes have only a certain # of colours.




```
Warning message:
In RColorBrewer::brewer.pal(n, pal) :
  n too large, allowed maximum for palette Accent is 8
Returning the palette you asked for with that many colors
```

Combine coolness now!!

You can choose a different palette



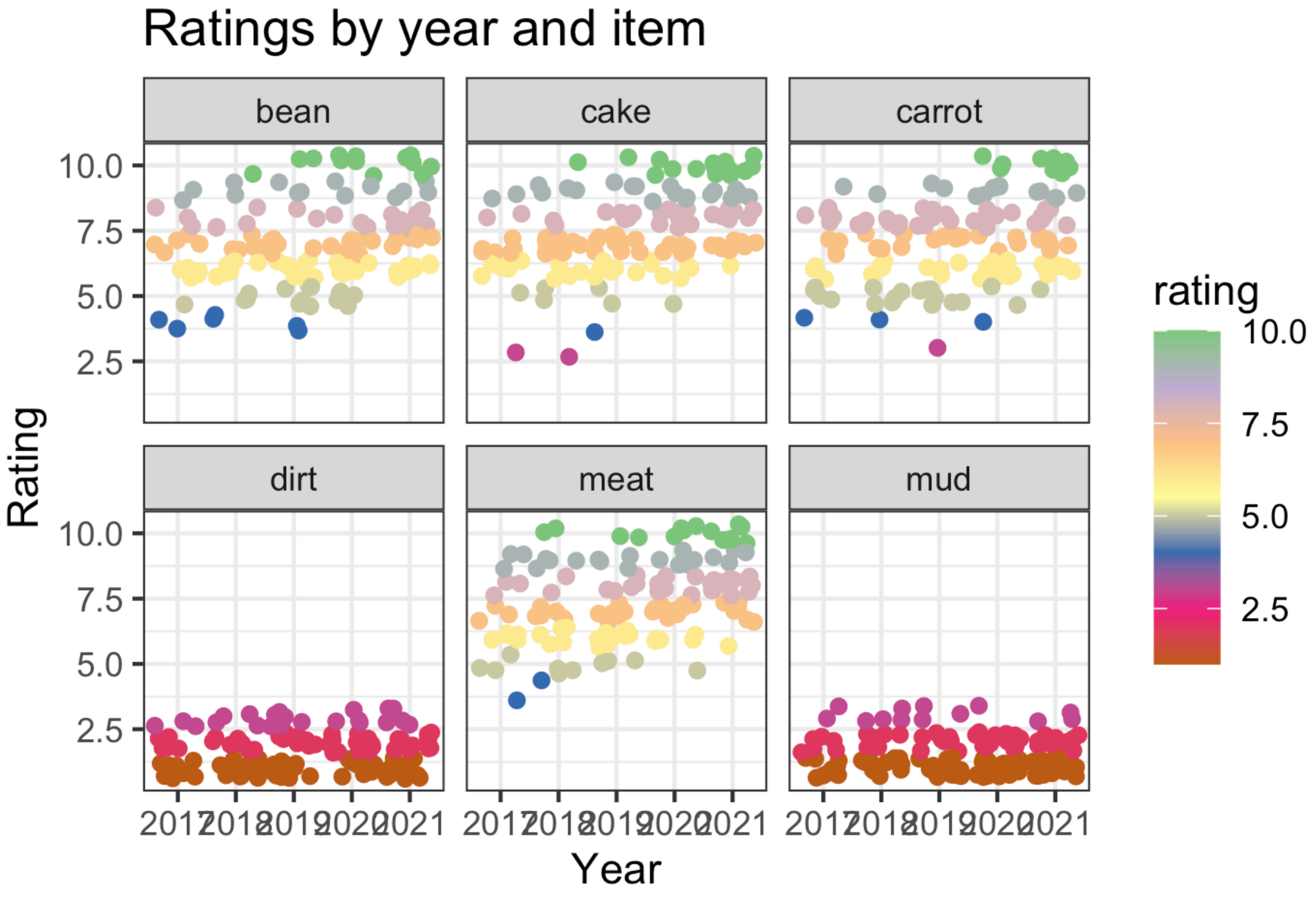

```
scale_fill_brewer(palette="Set3") +
```

Combine coolness now!!

… or if your variable is continuous,
use “distiller” instead to interpolate
between values

```
dl2 %>%
  ggplot(mapping = aes(x = year, y = rating,
                       fill=rating, colour=rating)) +
geom_jitter() +
scale_fill_distiller(palette="Accent") +
scale_colour_distiller(palette="Accent") +
facet_wrap(~question) +
theme_bw() +
labs(title = "Ratings by year and item",
  x = "Year",
  y = "Rating")
```

Combine coolness now!!



.. or you can use many of the other existing palettes out there besides RColorBrewer

See the `w4day2exercises.Rmd` file for the exercises!