MONASH University

MONASH BUSINESS SCHOOL

ETC3250: Data visualisation

Week 6, class 2

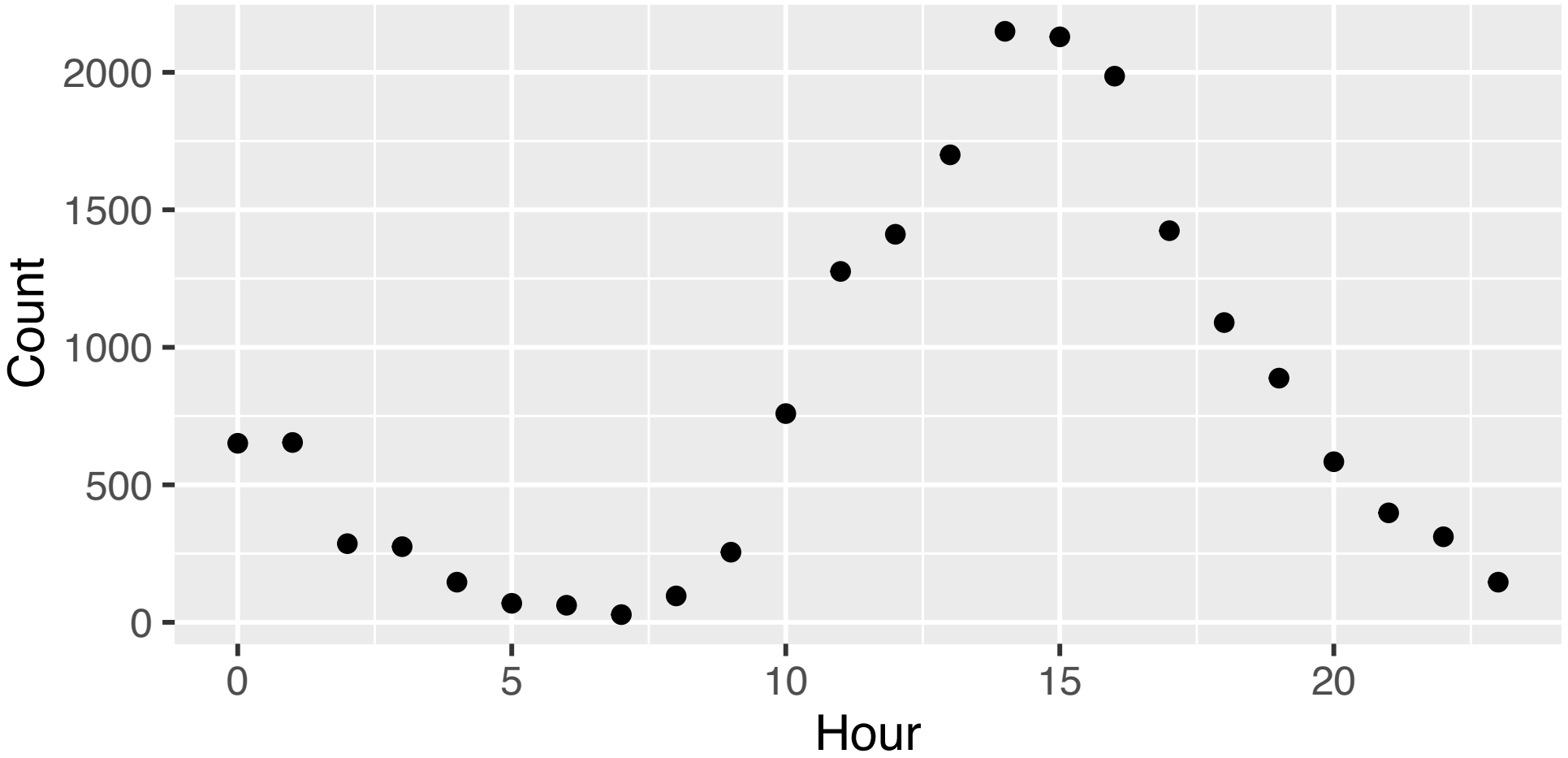
Professor Di Cook, Econometrics and Business Statistics

Data - Pedestrian Sensor

```
# Observations: 831,639
# Variables: 10
# $ DateTime   <chr> "01-JAN-2013 00:00", "01-JAN-2013 00:00"
# $ SensorID   <int> 4, 17, 18, 16, 2, 1, 13, 3, 9, 10, 12, 1
# $ SensorName <chr> "Town Hall (West)", "Collins Place (Soutl
# $ Counts     <int> 2992, 979, 413, 807, 785, 651, 599, 2562
# $ Date       <date> 2013-01-01, 2013-01-01, 2013-01-01, 201
# $ Year       <dbl> 2013, 2013, 2013, 2013, 2013, 2013, 2013
# $ Month      <ord> Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, J
# $ Day        <ord> Tues, Tues, Tues, Tues, Tues, Tues, Tues
# $ Hour       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
# $ Time       <dbl> 32184, 32184, 32184, 32184, 32184, 32184
```

Plotting points

```
ggplot(data=pedestrian, aes(x=Hour, y=Counts)) + geom_point()
```



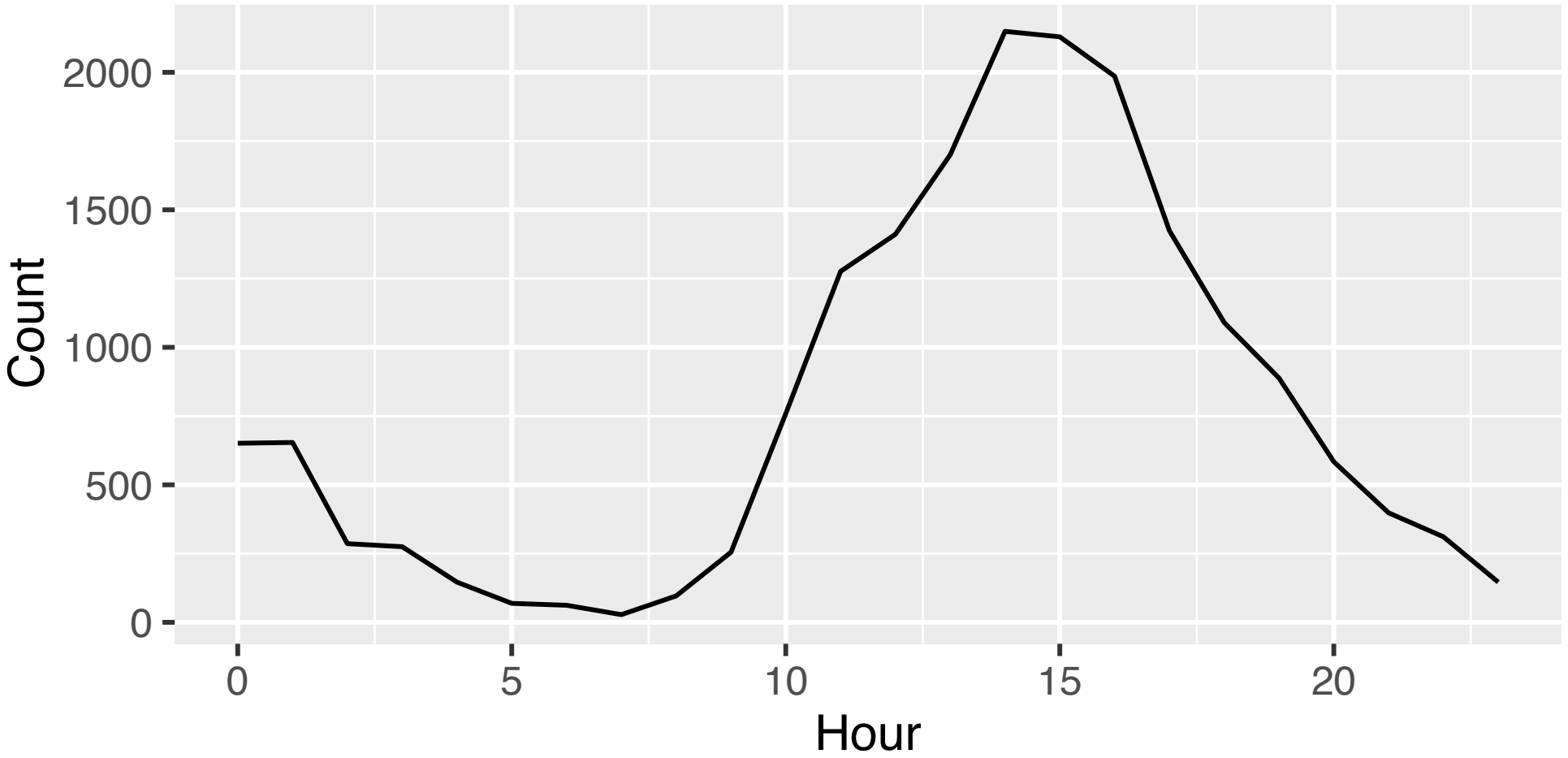
Your turn

How is the data mapped to graphical elements?

- data: _____
- aesthetics: _____
- geom: _____
- transformations: _____

Adding lines

```
ggplot(data=pedestrian, aes(x=Hour, y=Counts)) + geom_line()
```

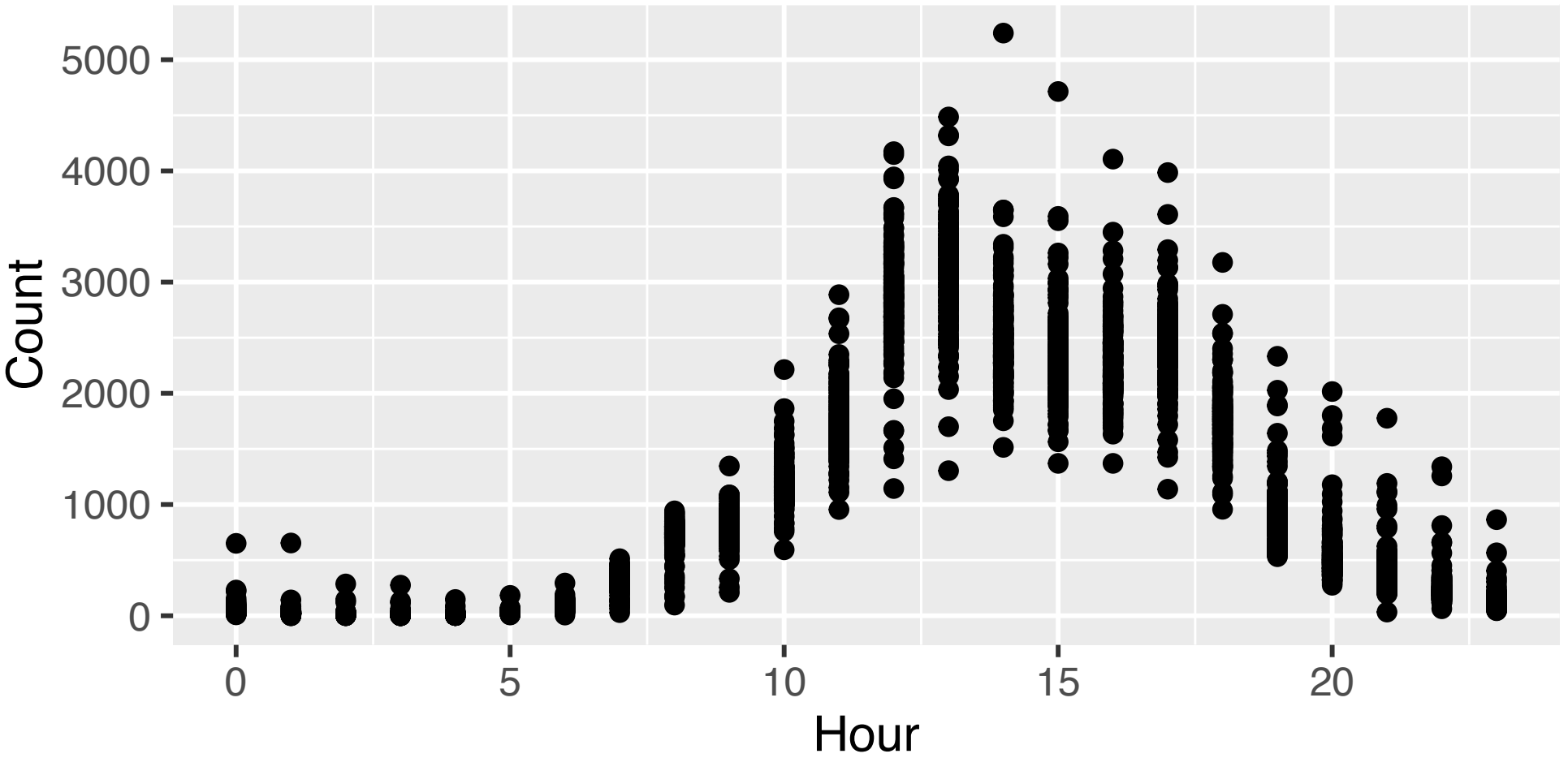


Your turn

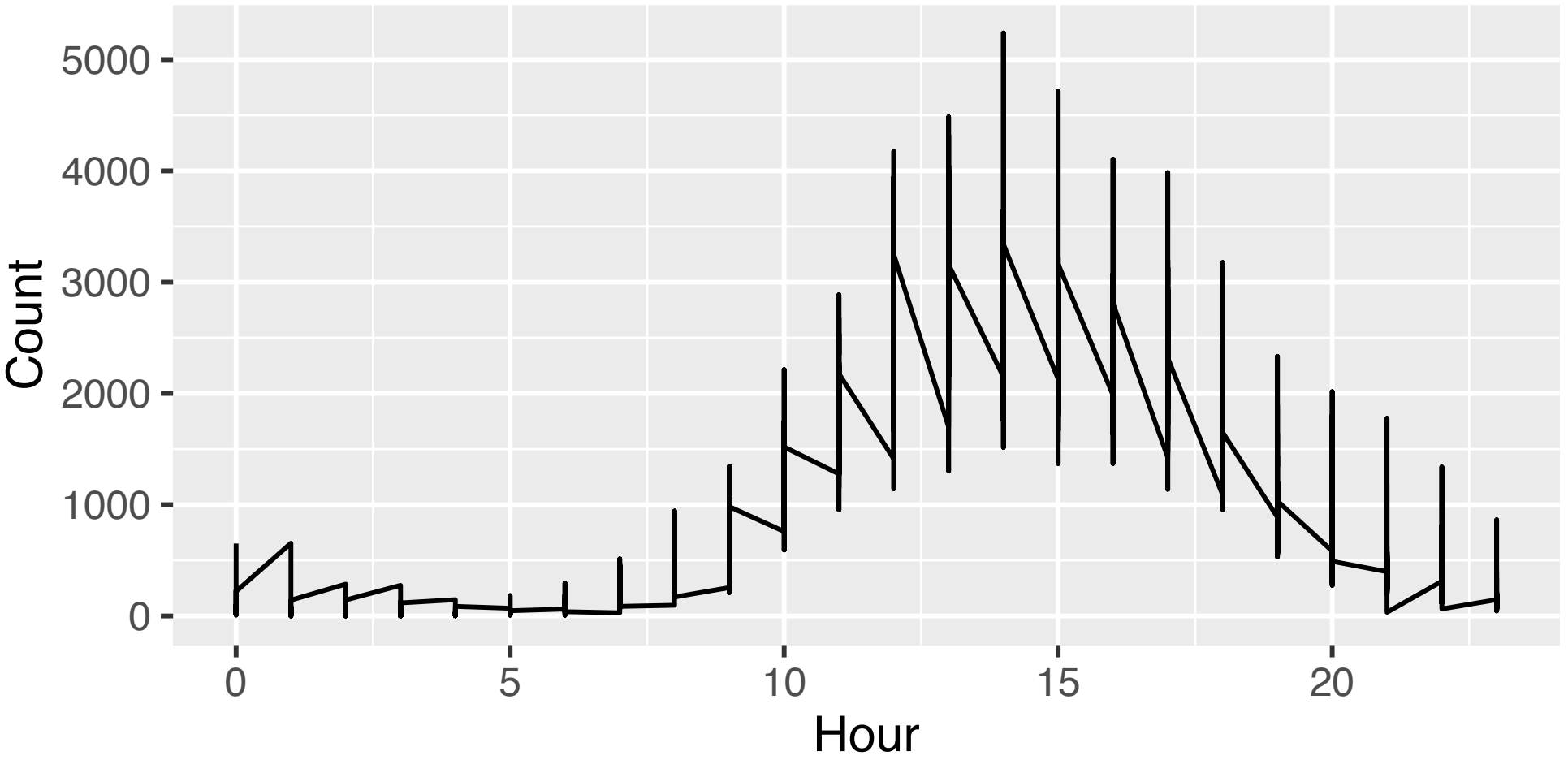
How is the data mapped to graphical elements?

- data: _____
- aesthetics: _____
- geom: _____
- transformations: _____

Multiple days



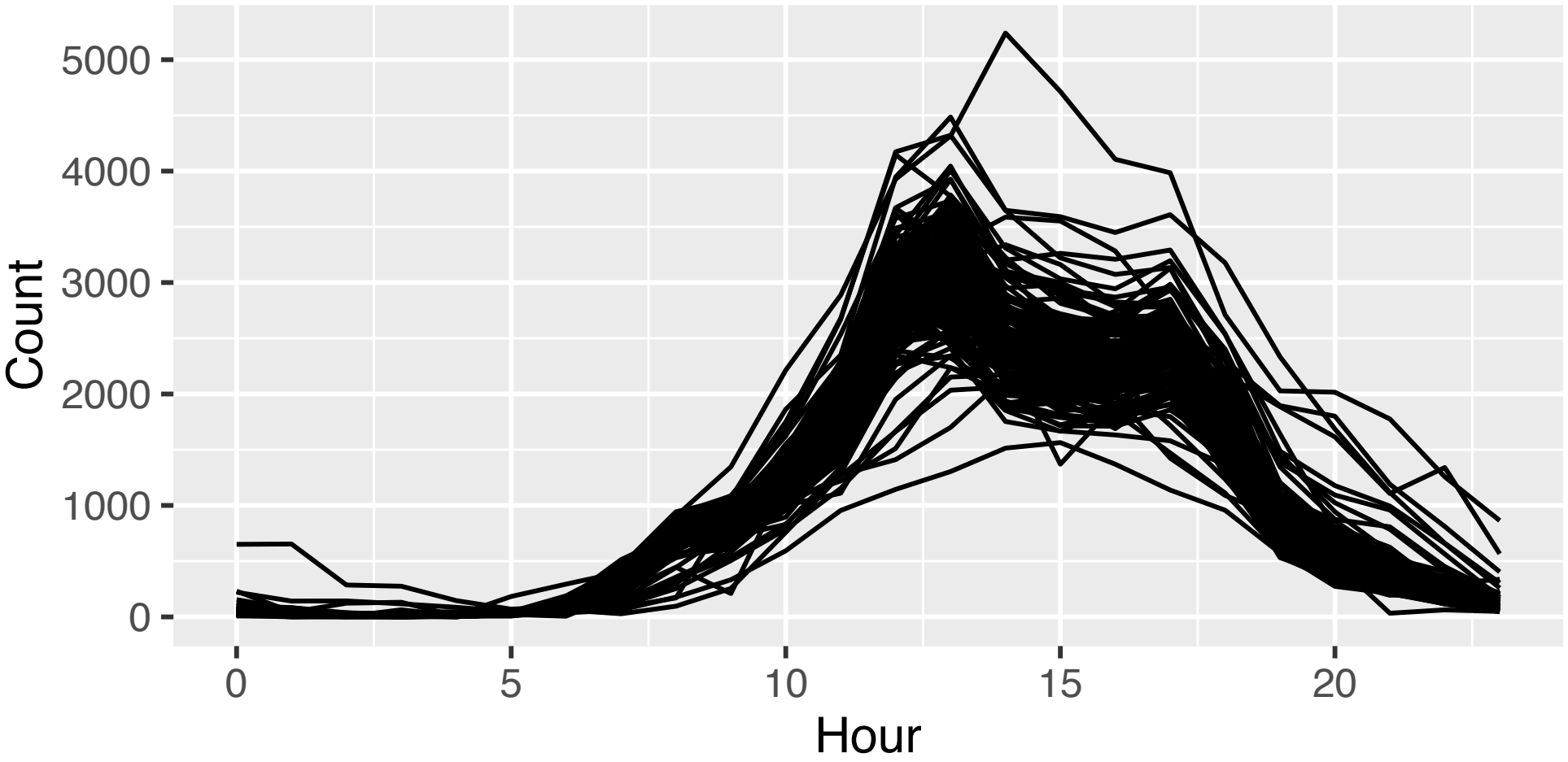
Use lines



That's not what I wanted ....

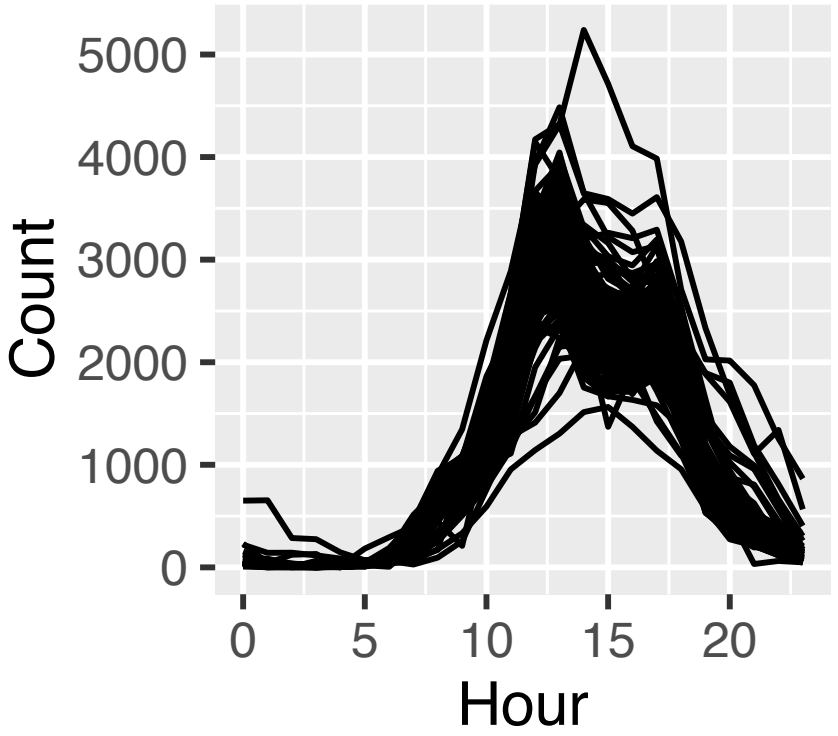
For each day grouped

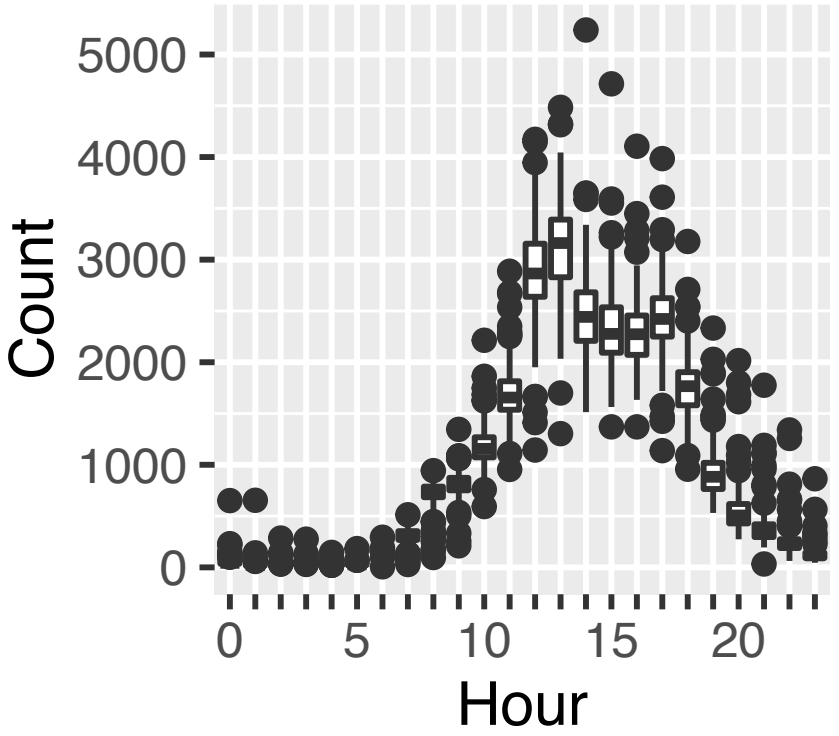
```
+ geom_line(aes(group=Date))
```



Which is better?

`+ geom_line(aes(group=Date))` vs `+ geom_boxplot()`





New example - Flying etiquette

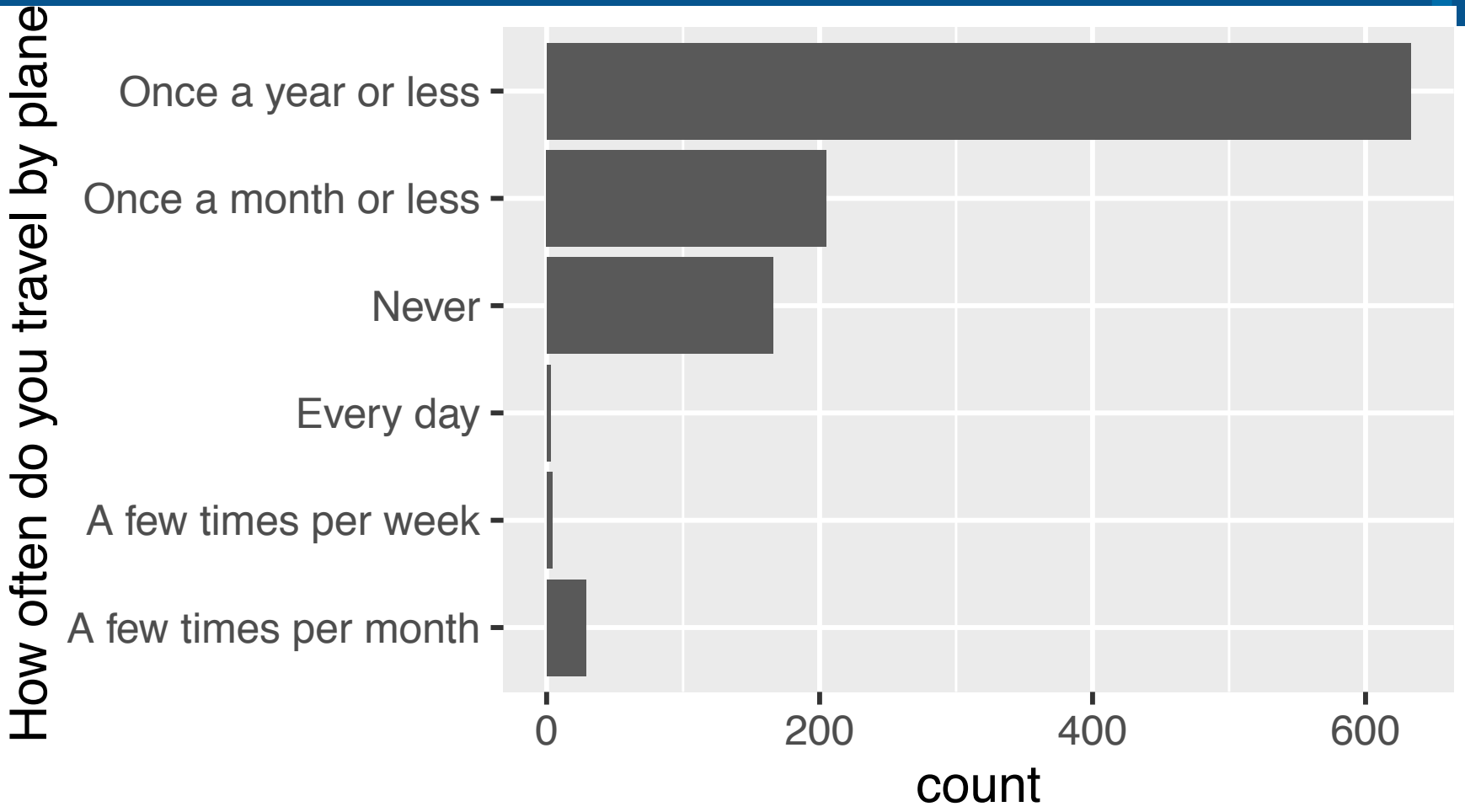
41% Of Fliers Think You're Rude If You Recline Your Seat

```
# Observations: 1,040
# Variables: 27
# $ RespondentID
# $ How often do you travel by plane?
# $ Do you ever recline your seat when you fly?
# $ How tall are you?
# $ Do you have any children under 18?
# $ In a row of three seats, who should get to use the two arm
# $ In a row of two seats, who should get to use the middle ar
# $ Who should have control over the window shade?
# $ Is itrude to move to an unsold seat on a plane?
# $ Generally speaking, is it rude to say more than a few word
# $ On a 6 hour flight from NYC to LA, how many times is it ac
# $ Under normal circumstances, does a person who reclines the
# $ Is itrude to recline your seat on a plane?
# $ Given the opportunity, would you eliminate the possibilit
```

Variables

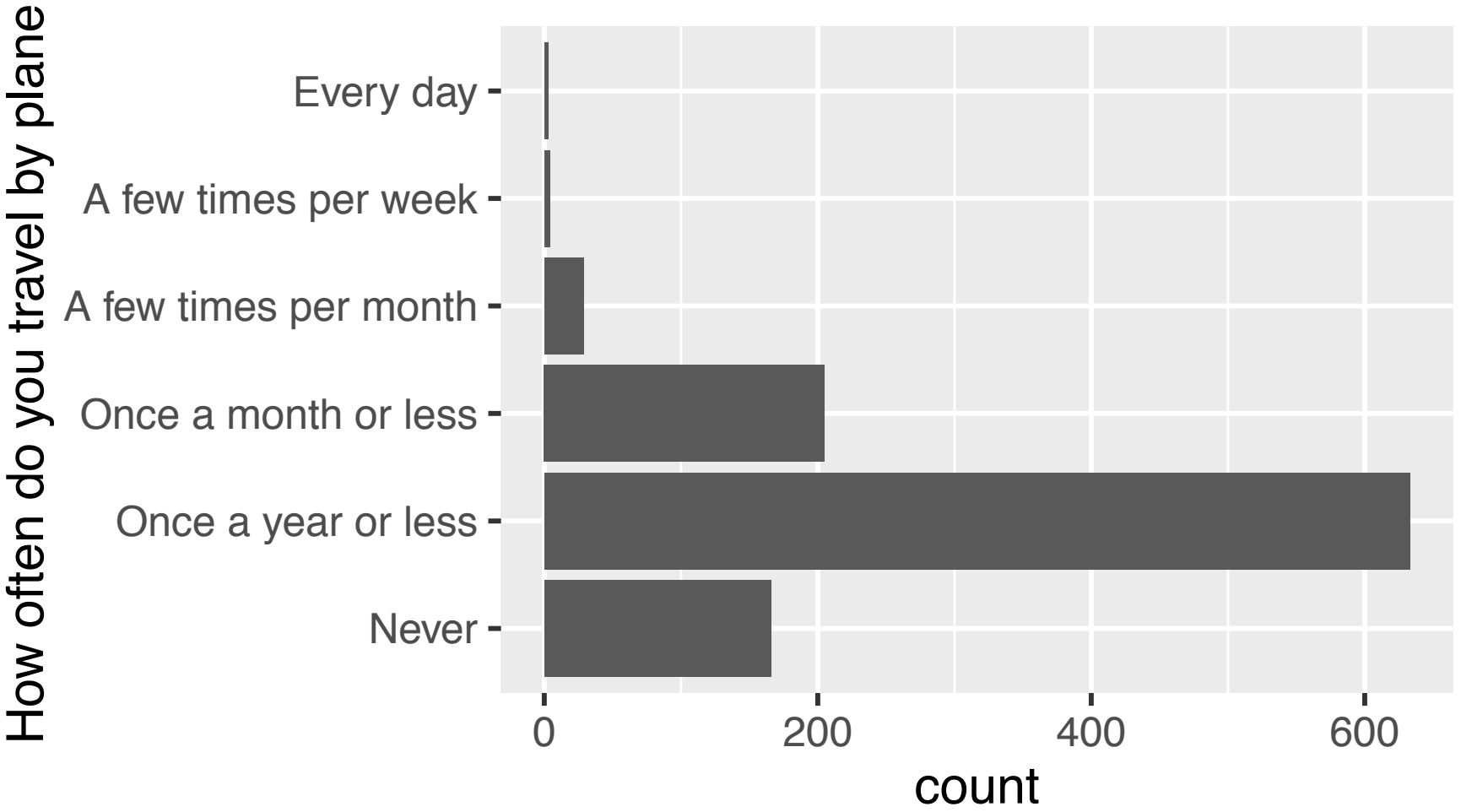
- Mix of categorical and quantiative variables.
- What mappings are appropriate?
- Area for counts of categories, side-by-side boxplots for mixed pair.

Support



Categories are not sorted

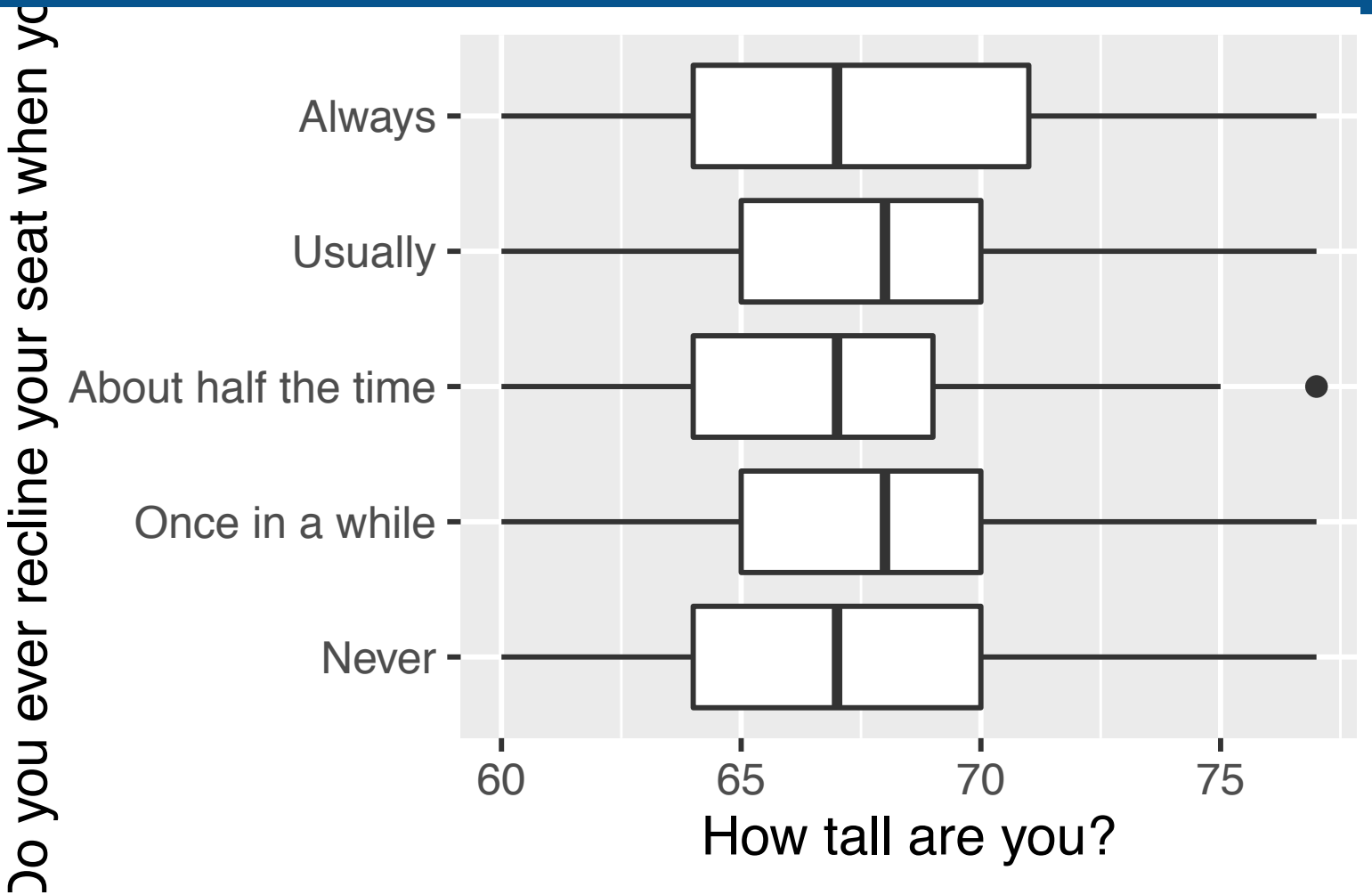
Sorted categories



Filter data

```
fly_sub <- fly %>%
  filter(`How often do you travel by plane?` %in%
    c("Once a year or less","Once a month or less")) %>%
  filter(!is.na(`Do you ever recline your seat when you fly?`)
  filter(!is.na(Age)) %>% filter(!is.na(Gender))
```

Recline by height



Cheat sheet

Take a look at the ggplot2 Cheat sheet

Your turn

How many geoms are available in ggplot2? What is `geom_rug`?

Your turn

What is the difference between `colour` and `fill`?

Your turn

What does `coord_fixed()` do? What is the difference between this and using `theme(aspect.ratio=...)` or `coord_equal()`?

Your turn

What are scales? How many numeric transformation scales are there?

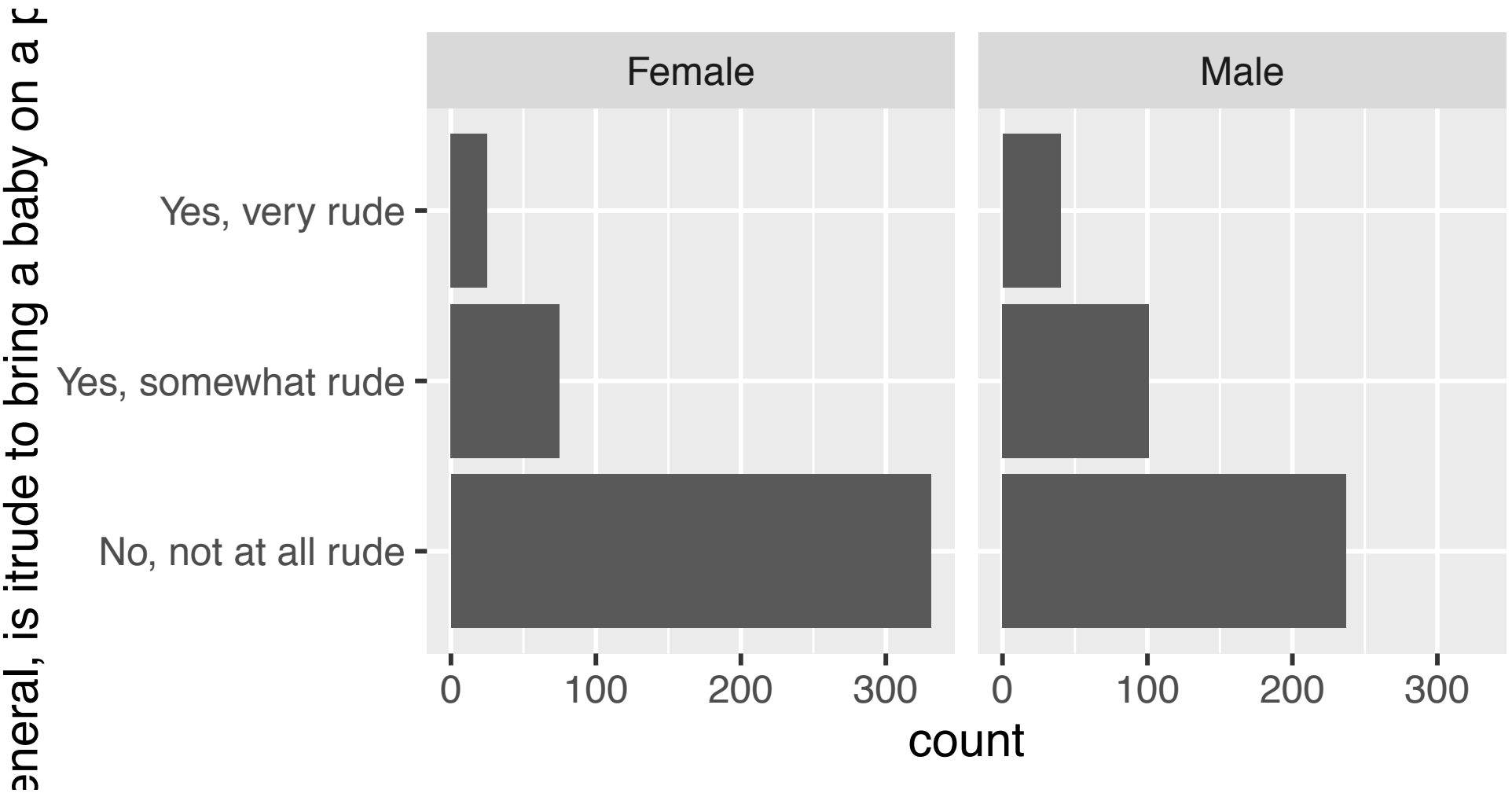
Your turn

What are position adjustments? When would they be used?

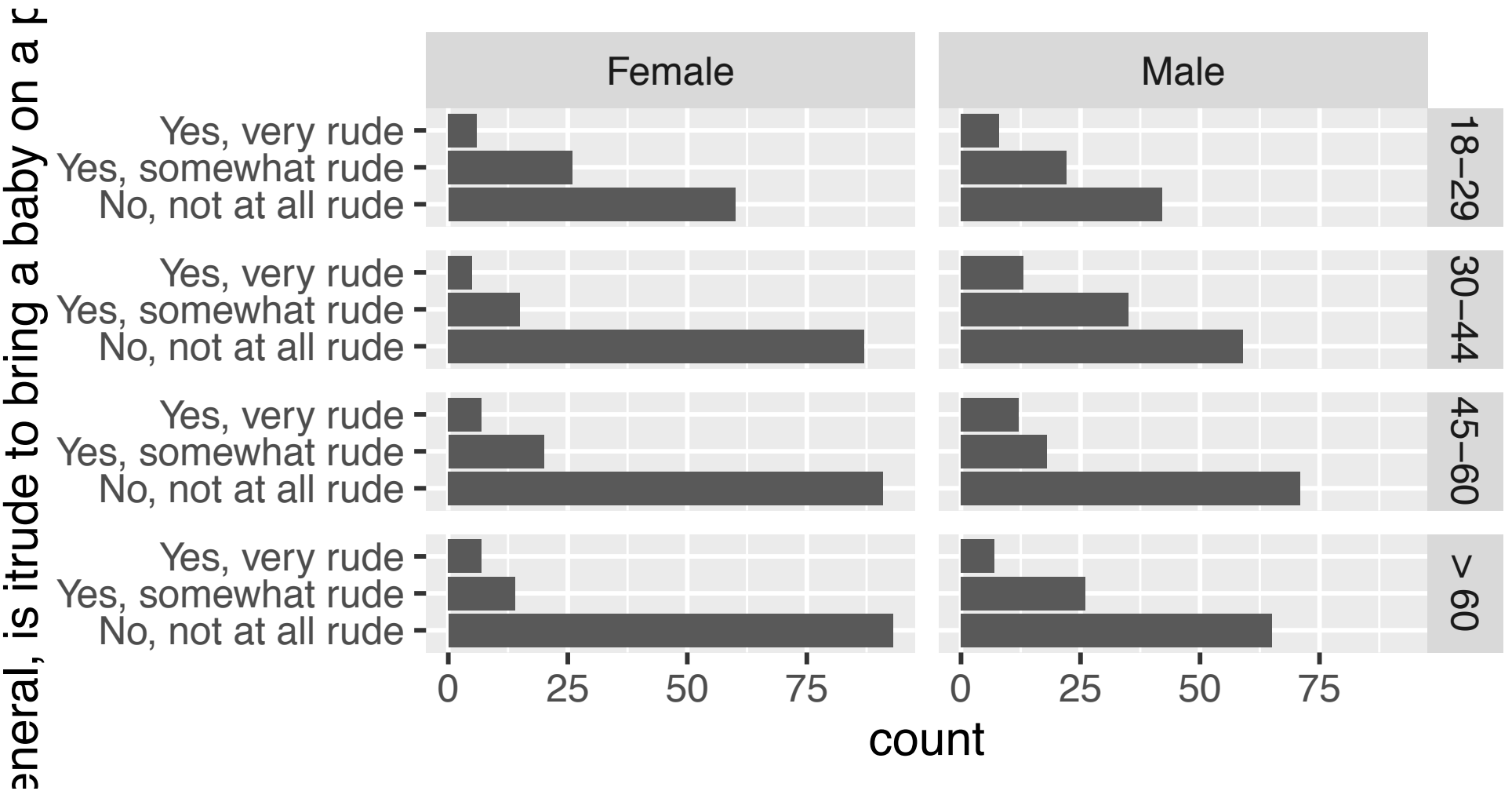
Your turn

- What type of plot do you need to make to explore the relationship between `Do you ever recline your seat when you fly?` and `Is it rude to recline your seat on a plane?`
- Use your cheat sheet to work out how to make it.

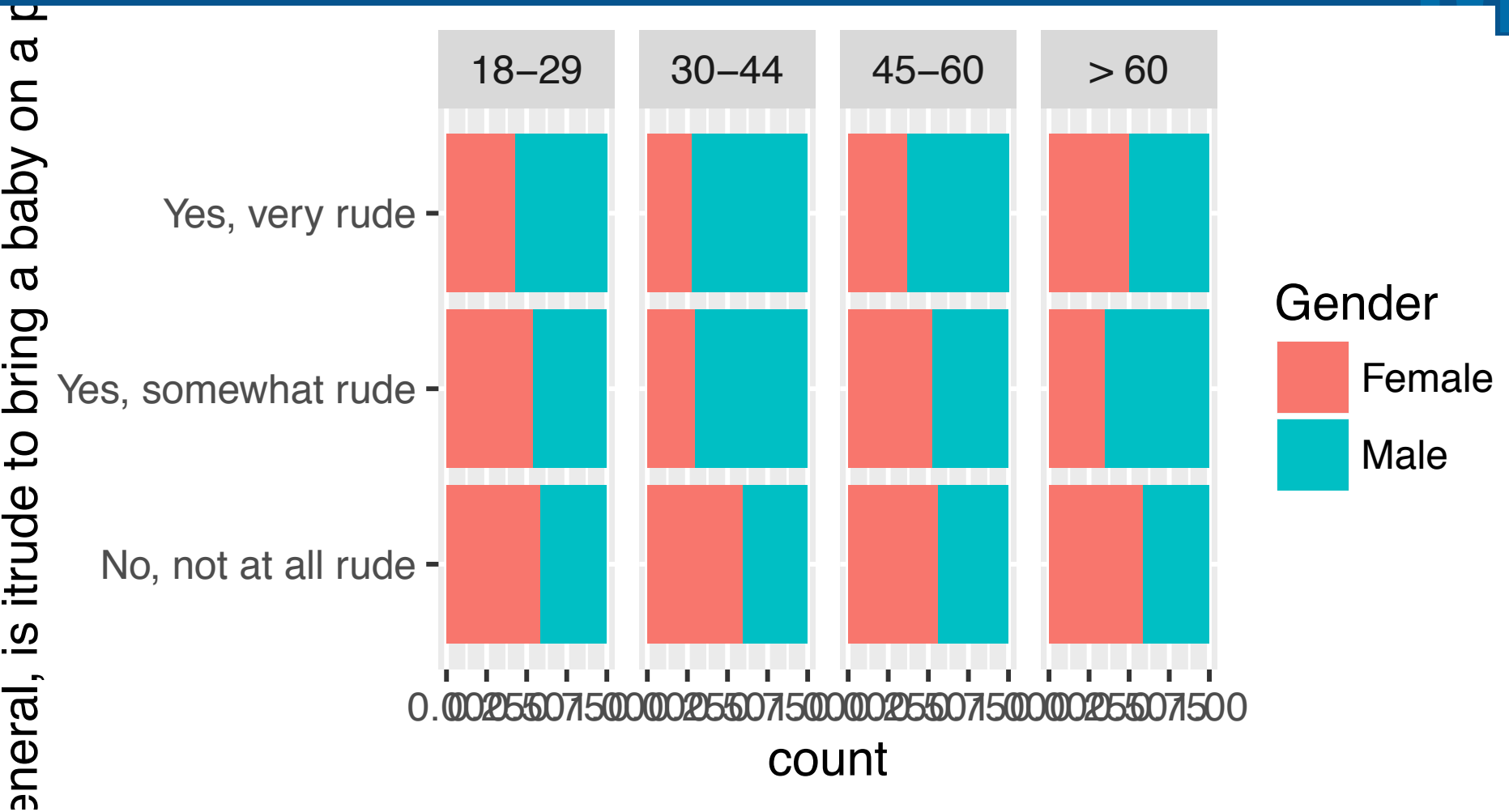
Facets


eneral, is itrude to bring a baby on a p
Female
Male
Yes, very rude
Yes, somewhat rude
No, not at all rude
0
100
200
300
count

Facets

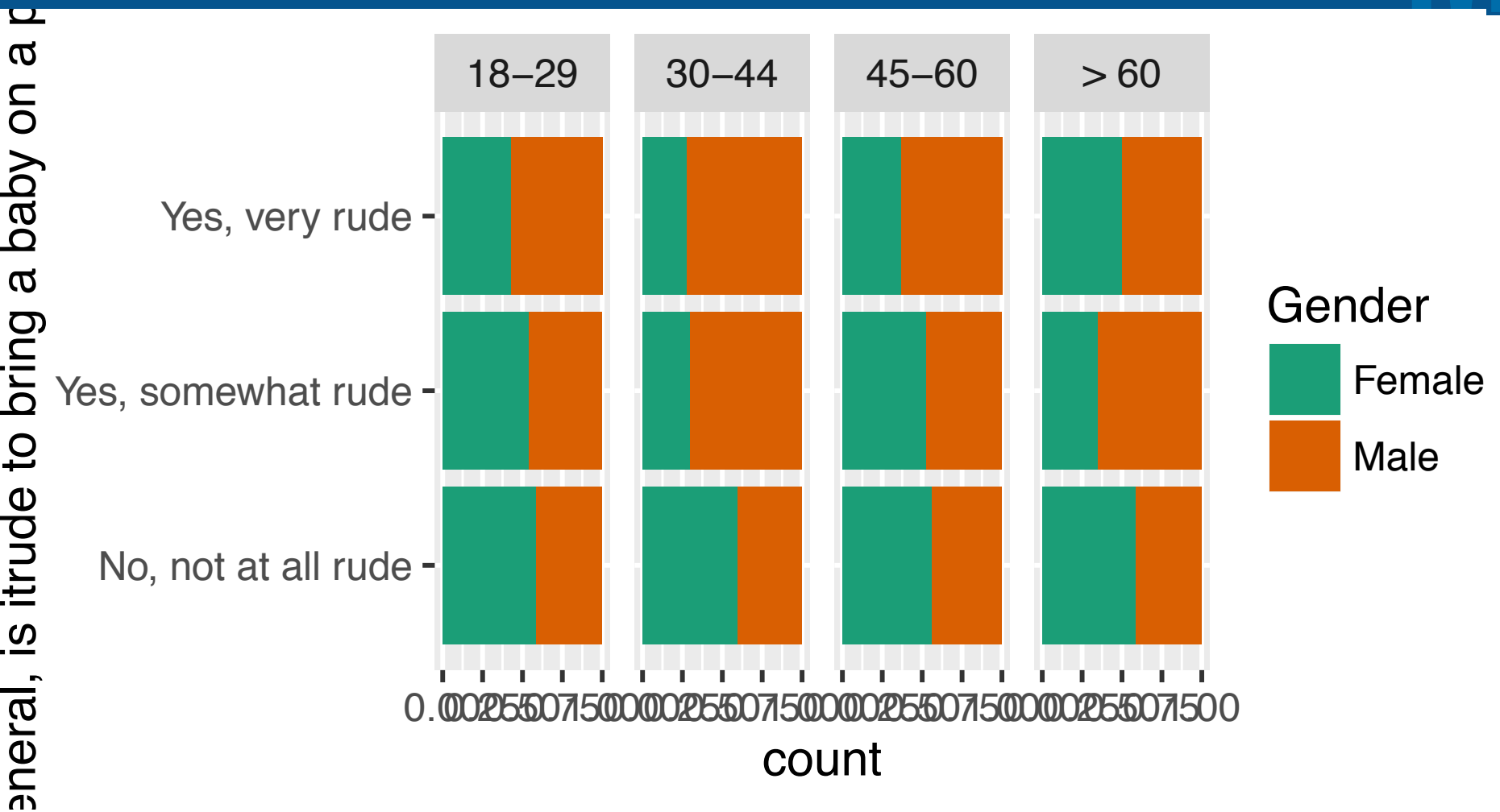

eneral, is itrude to bring a baby on a p
Female
Male
18–29
30–44
45–60
> 60
Yes, very rude
Yes, somewhat rude
No, not at all rude
0
25
50
75
count

Color palettes - default



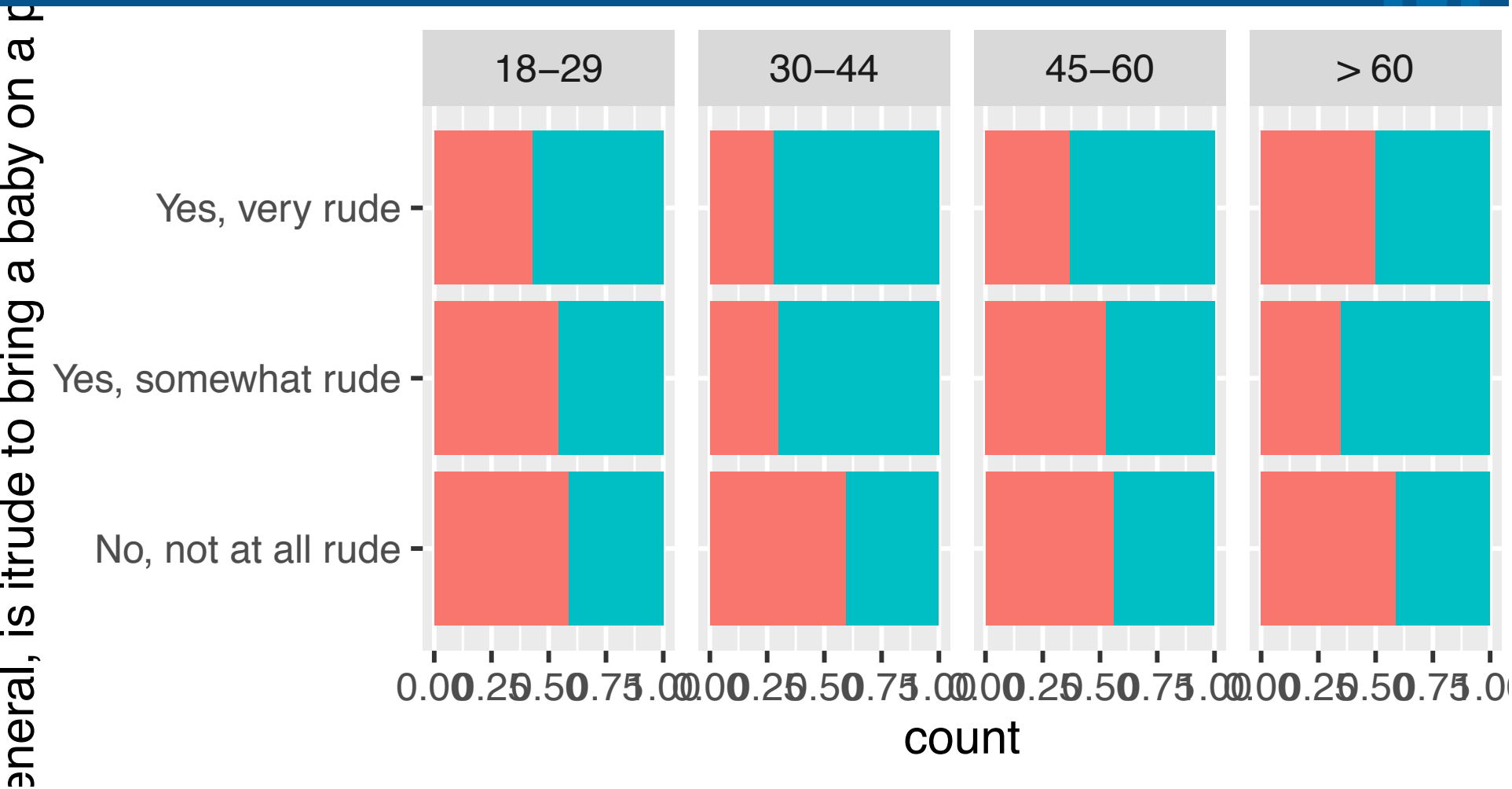
What do we learn?

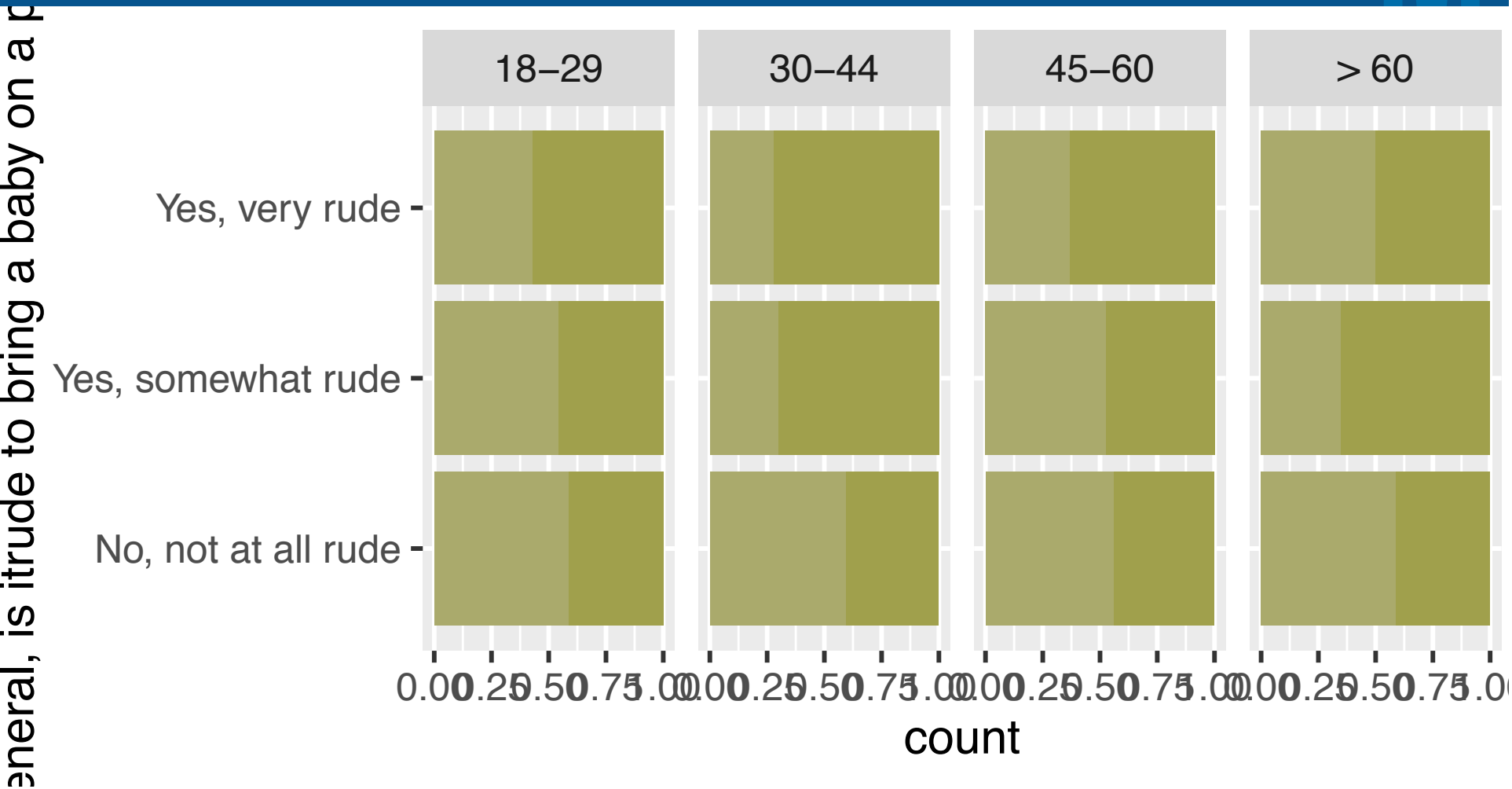
Color palettes - brewer



Color blind-proofing

```
library(scales)
library(dichromat)
clrs <- hue_pal()(3)
p + theme(legend.position = "none")
clrs <- dichromat(hue_pal()(3))
p + scale_fill_manual("", values=clrs) + theme(legend.positio
```


eneral, is itrude to bring a baby on a p
18–29
30–44
45–60
> 60
Yes, very rude
Yes, somewhat rude
No, not at all rude
0.00
0.25
0.50
0.75
1.00
count


eneral, is itrude to bring a baby on a p
18–29
30–44
45–60
> 60
Yes, very rude
Yes, somewhat rude
No, not at all rude
0.00
0.25
0.50
0.75
1.00
count

Perceptual principles

- Hierarchy of mappings: (first) position along an axis - (last) color (Cleveland, 1984; Heer and Bostock, 2009)
- Pre-attentive: Some elements are noticed before you even realise it.
- Color: (pre-attentive) palettes - qualitative, sequential, diverging.
- Proximity: Place elements for primary comparison close together.
- Change blindness: When focus is interrupted differences may not be noticed.

Hierarchy of mappings

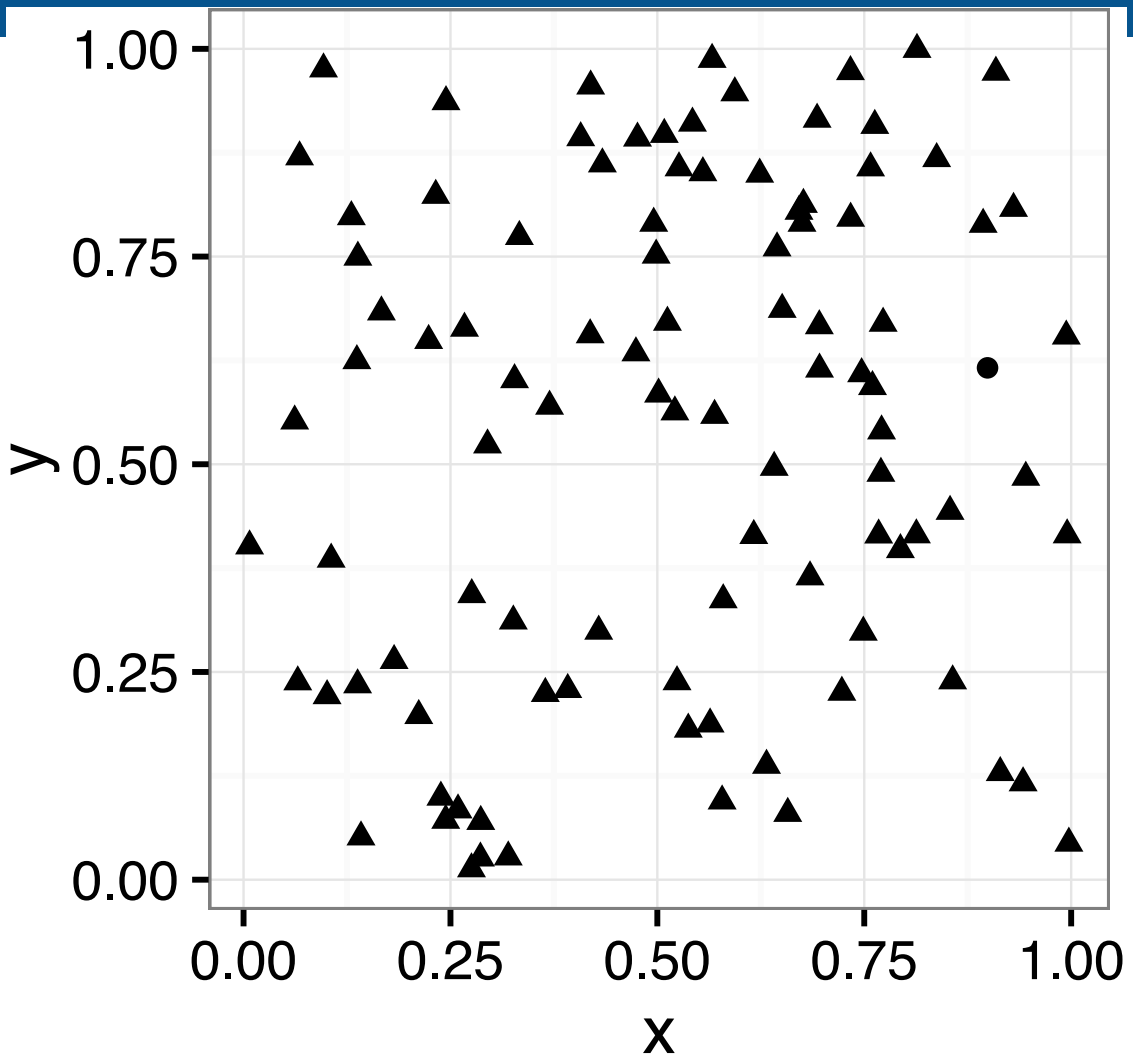
1. Position - common scale (BEST)
2. Position - nonaligned scale
3. Length, direction, angle
4. Area
5. Volume, curvature
6. Shading, color (WORST)

See my blog post for a re-do with crowd-sourcing Aug 3, 2016.

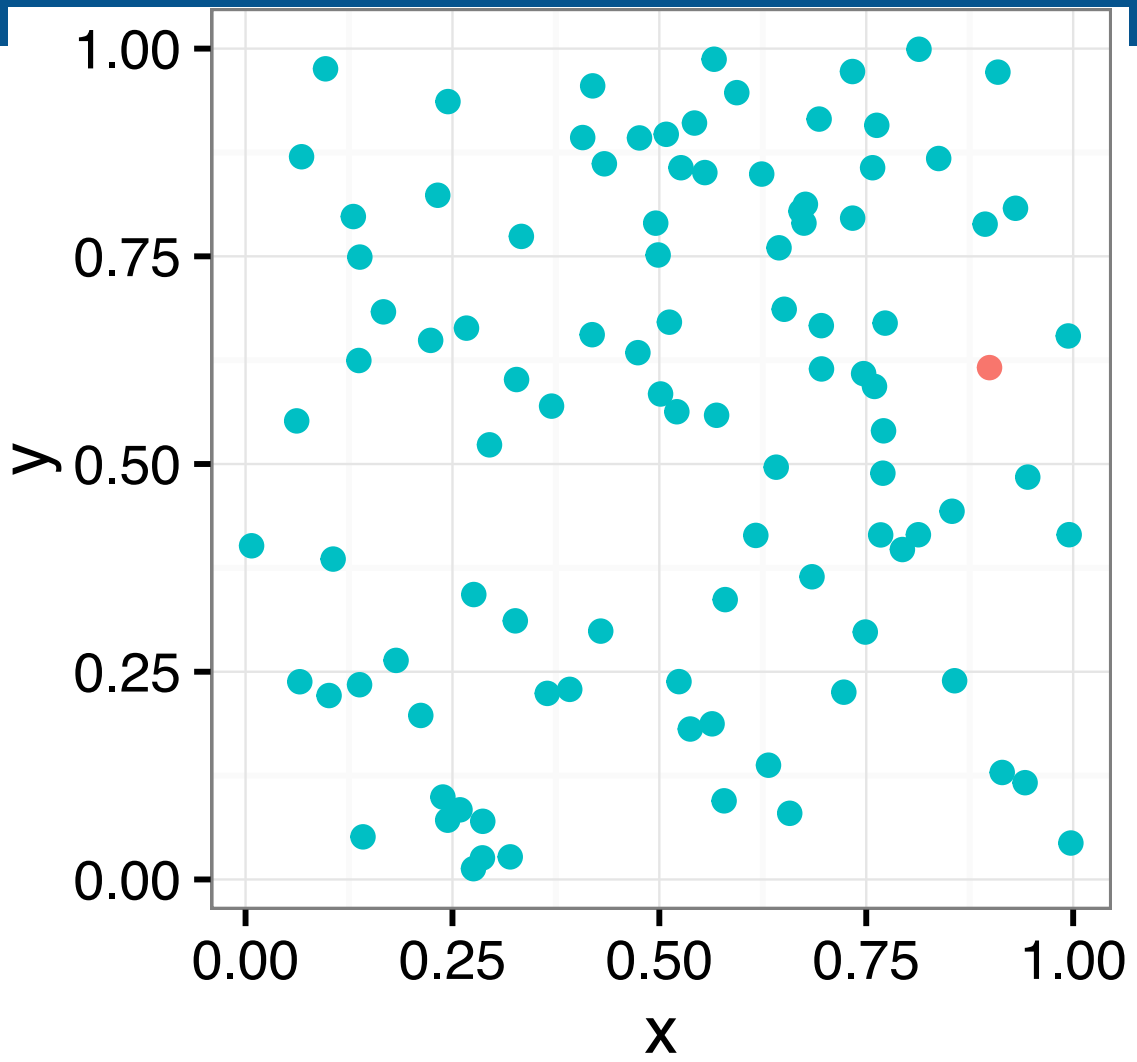
Pre-attentive

"A preattentive visual property is one which is processed in spatial memory without our conscious action. In essence it takes less than 500 milliseconds for the eye and the brain to process a preattentive property of any image."" Source: Interaction Design Foundation

Can you find the odd one out?

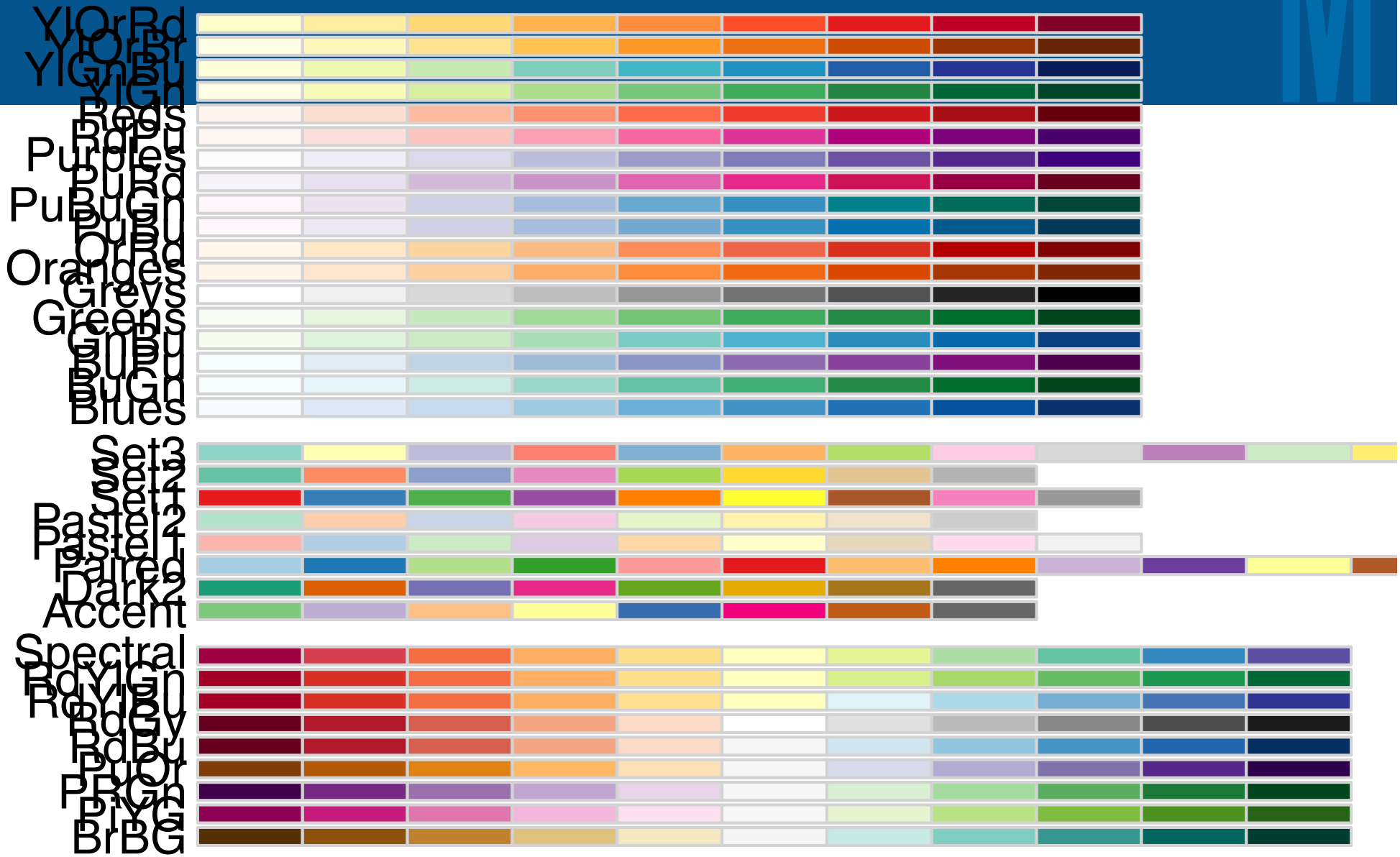


Is it easier now?



Color palettes

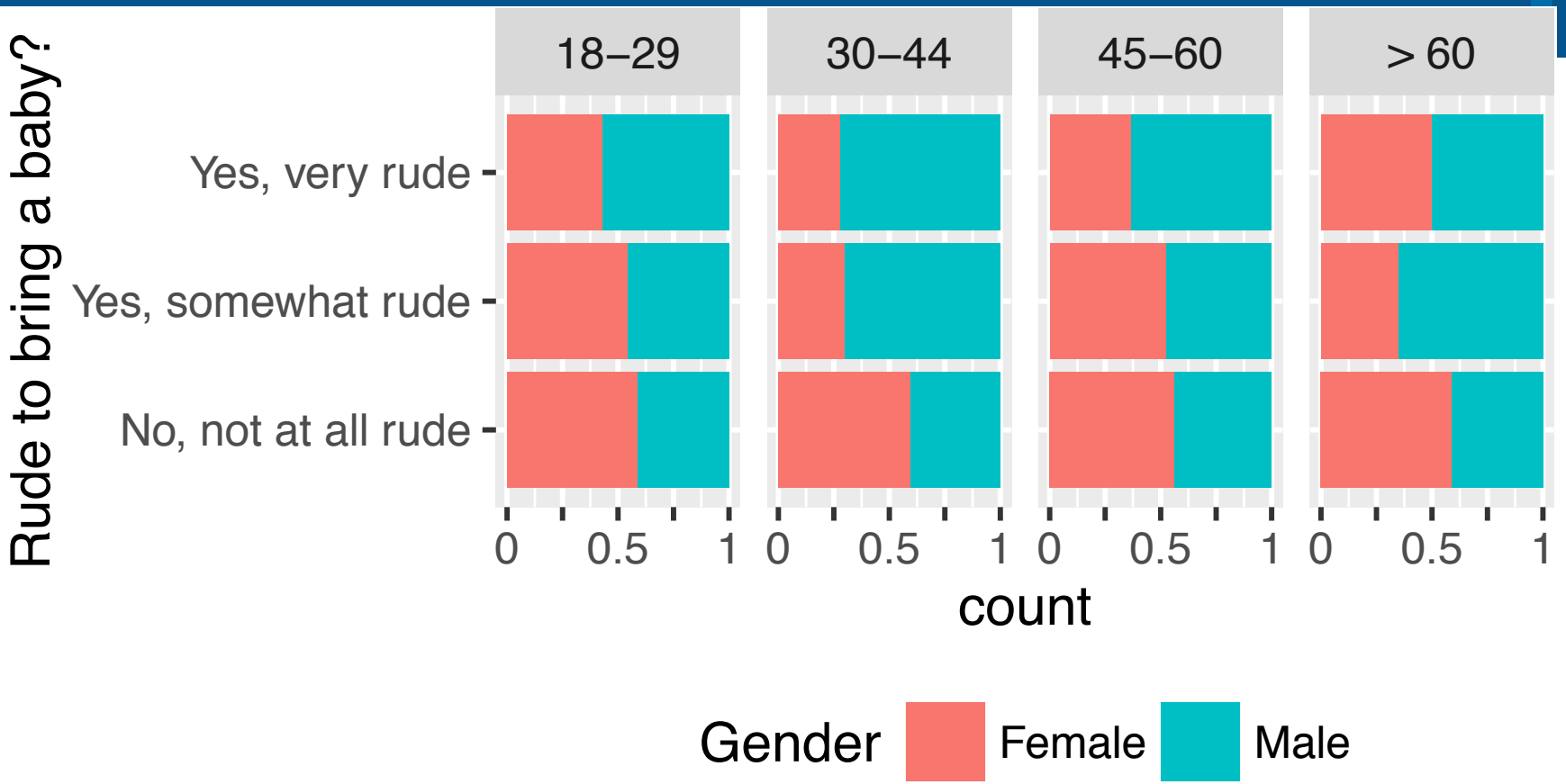
- Qualitative: categorical variables
- Sequential: low to high numeric values
- Diverging: negative to positive values


YlOrRd
YlOrBr
YlGnBu
YlGn
Reds
RdPu
Purples
PuBuGn
PuBu
Oranges
Greys
Greens
GnBu
BuPu
BuGn
Blues
Set3
Set2
Set1
Pastel2
Pastel1
Paired
Dark2
Accent
Spectral
RdYlGn
RdYlBu
RdGy
RdBu
PuOr
PRGn
PiYG
BrBG

Proximity

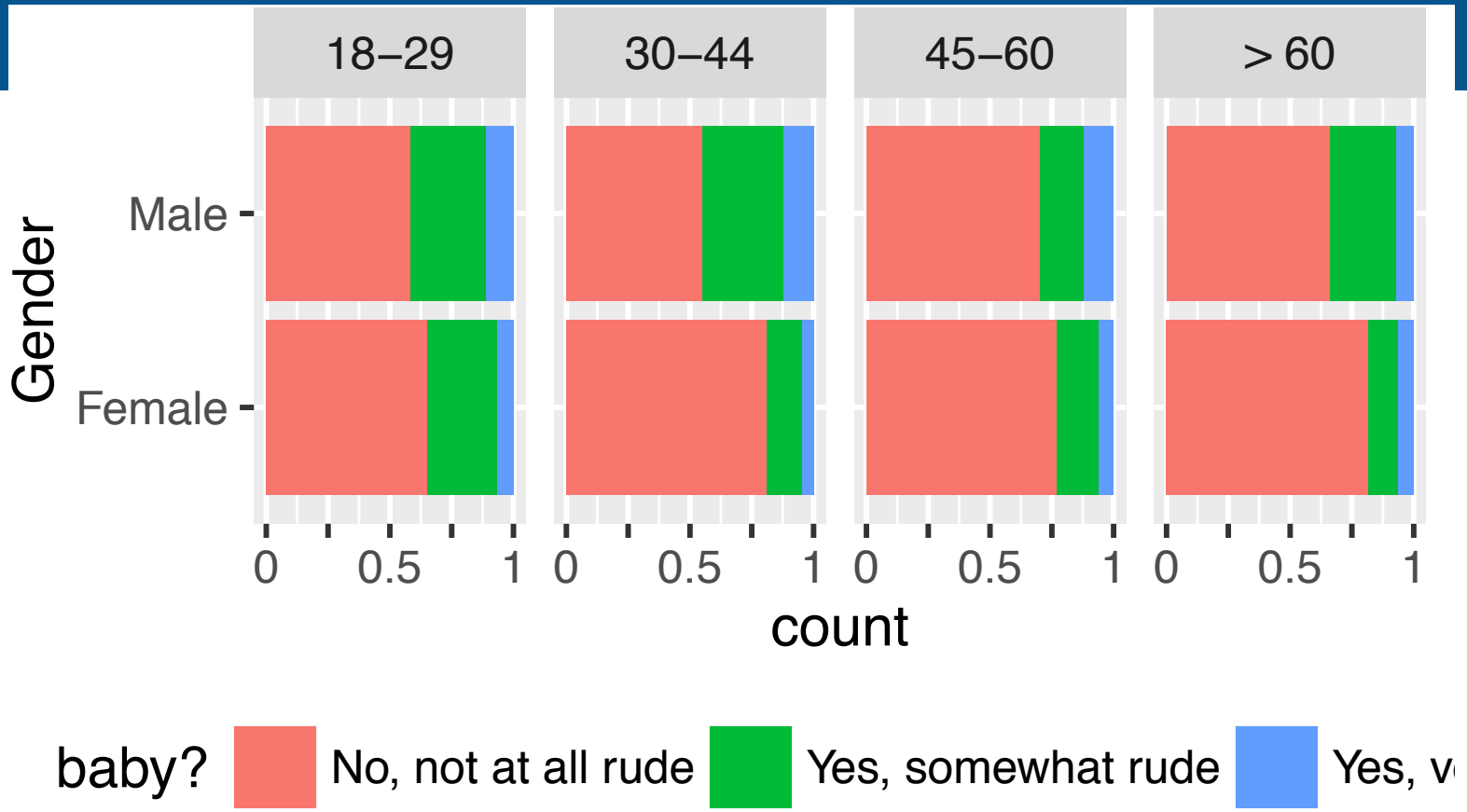
```
ggplot(fly_sub,
       aes(x=`In general, is itrude to bring a baby on a plane
                    fill=Gender)) +
  geom_bar(position="fill") + coord_flip() + facet_wrap(~Age,
  scale_y_continuous(breaks=seq(0,1,0.25),
                     labels=c("0","","0.5","","1")) +
  xlab("Rude to bring a baby?") +
  theme(legend.position="bottom")
```

With this arrangement we can see proportion of gender within each rudeness category, and compare these across age groups. How could we arrange this differently?


Rude to bring a baby?
18–29
30–44
45–60
> 60
Yes, very rude
Yes, somewhat rude
No, not at all rude
0
0.5
1
count
Gender
Female
Male

Proximity

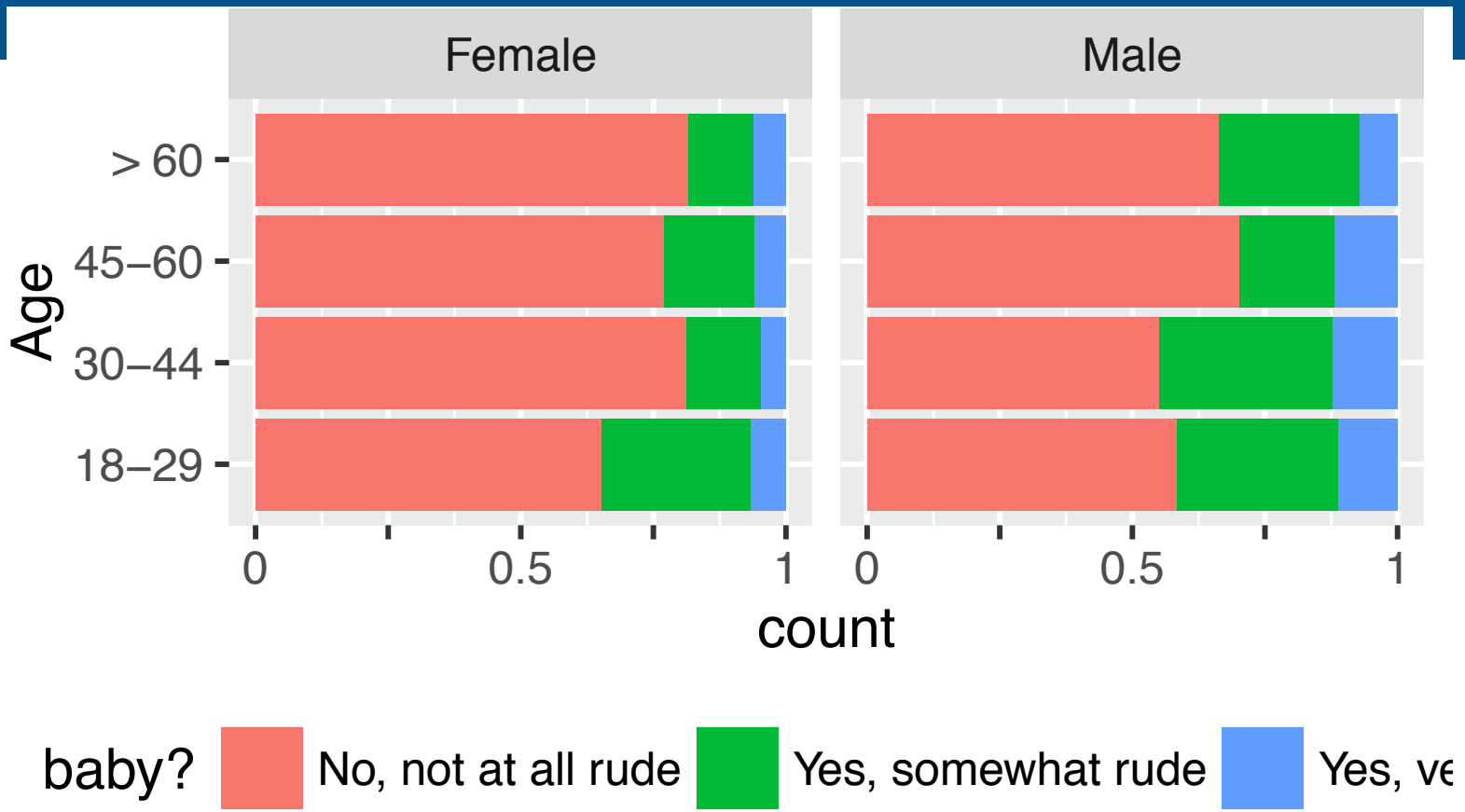
```
ggplot(fly_sub, aes(x=Gender,
   fill=`In general, is itrude to bring a baby on a plane?`))
  geom_bar(position="fill") + coord_flip() + facet_wrap(~Age,
  scale_y_continuous(breaks=seq(0,1,0.25),
                     labels=c("0","","0.5","","1")) +
  scale_fill_discrete("baby?") +
  theme(legend.position="bottom")
```



What is different about the comparison now?

Another arrangement

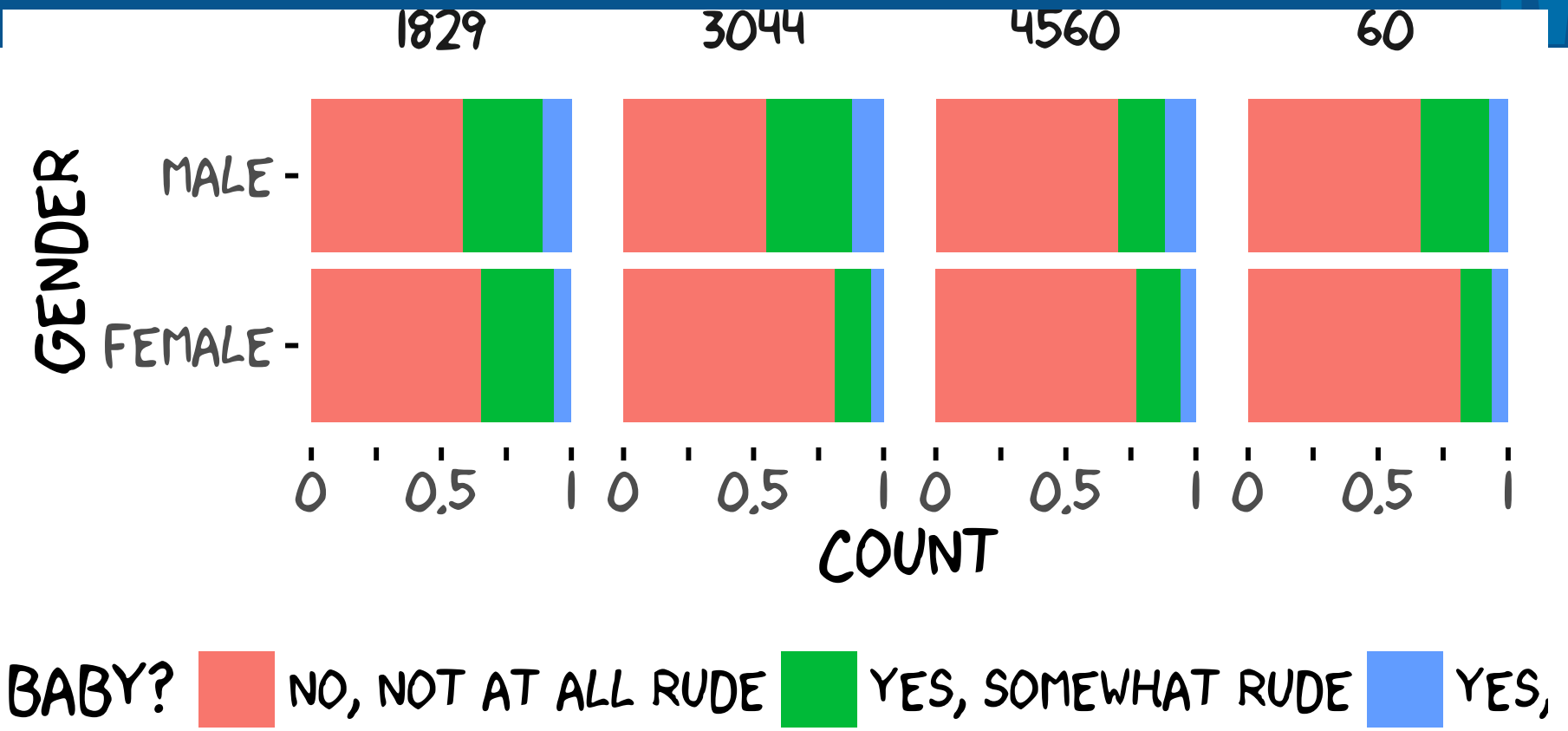
```
ggplot(fly_sub, aes(x=Age,
           fill=`In general, is itrude to bring a baby on a pla
  geom_bar(position="fill") + coord_flip() +
  scale_y_continuous(breaks=seq(0,1,0.25),
                     labels=c("0","","0.5","","1")) +
  scale_fill_discrete("baby?") +
  facet_wrap(~Gender, ncol=5) + theme(legend.position="bottom"
```


Female
Male
> 60
45–60
30–44
18–29
Age
0
0.5
1
count
baby?
No, not at all rude
Yes, somewhat rude
Yes, ve

Themes

The `ggthemes` package has many different styles for the plots. Other packages such as `xkcd`, `skittles`, `wes anderson`, `beyonce`, ...
See the vignette for instructions on installing the xkcd font.

```
library(xkcd)
ggplot(fly_sub, aes(x=Gender,
   fill=`In general, is itrude to bring a baby on a plane?`))
  geom_bar(position="fill") + coord_flip() +
  facet_wrap(~Age, ncol=5) +
  scale_y_continuous(breaks=seq(0,1,0.25),
                     labels=c("0","","0.5","","1")) +
  scale_fill_discrete("baby?") +
  theme_xkcd() + theme(legend.position="bottom")
```


1829
3044
4560
60
GENDER
MALE
FEMALE
0
0.5
1
COUNT
BABY?
NO, NOT AT ALL RUDE
YES, SOMEWHAT RUDE
YES,

Resources

- Cheat sheet
- ggplot2: Elegant Graphics for Data Analysis, Hadley Wickham, web site
- R Graphics Cookbook, Winston Chang
- Naomi Robbins, Creating More Effective Graphs
- Antony Unwin, Graphical Data Analysis with R

Share and share alike