

r4ds Chapter 3

Smiti Kaul

Jan 22 - Feb 1, 2018

Modify SWDA data

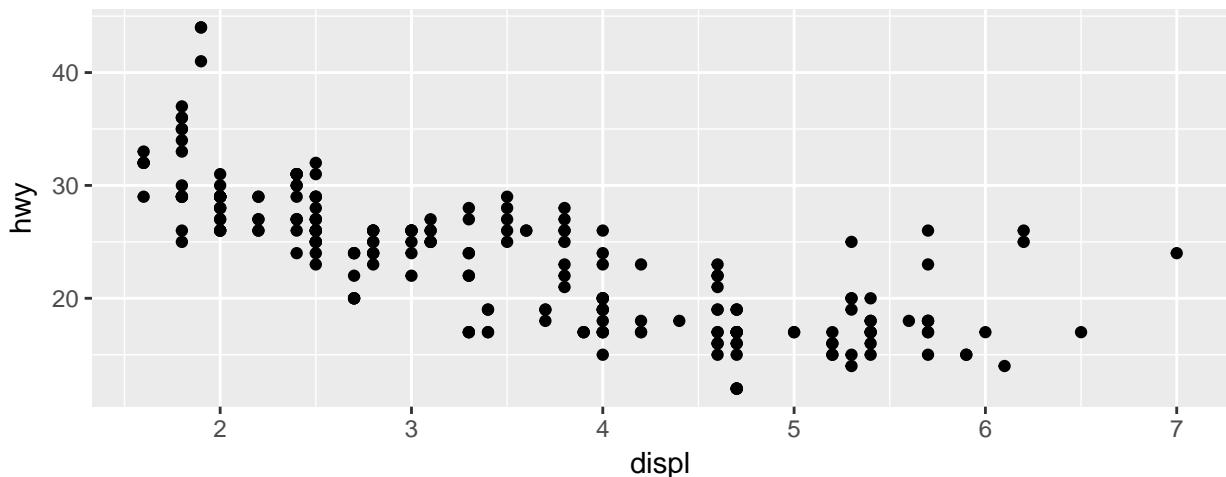
```
df <- swda
df$sex[df$sex == 'MALE'] <- 0
df$sex[df$sex == 'FEMALE'] <- 1

# sort(unique(df$dialect_area))
df$dialect_area[df$dialect_area == 'MIXED'] <- 0
df$dialect_area[df$dialect_area == 'NEW ENGLAND'] <- 1
df$dialect_area[df$dialect_area == 'NORTH MIDLAND'] <- 2
df$dialect_area[df$dialect_area == 'NORTHERN'] <- 3
df$dialect_area[df$dialect_area == 'NYC'] <- 4
df$dialect_area[df$dialect_area == 'SOUTH MIDLAND'] <- 5
df$dialect_area[df$dialect_area == 'SOUTHERN'] <- 6
df$dialect_area[df$dialect_area == 'UNK'] <- 7
df$dialect_area[df$dialect_area == 'WESTERN'] <- 8

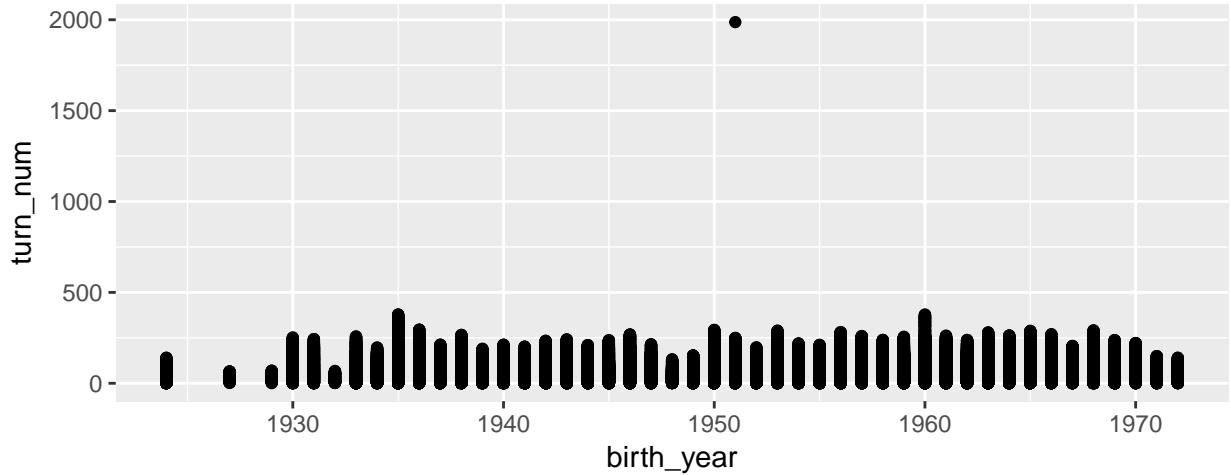
df$sex <- as.numeric(as.character(df$sex))
df$dialect_area <- as.numeric(as.character(df$dialect_area))
df$utterance_num <- as.numeric(as.character(df$utterance_num))
df$turn_num <- as.numeric(as.character(df$turn_num))
```

3.2: First steps

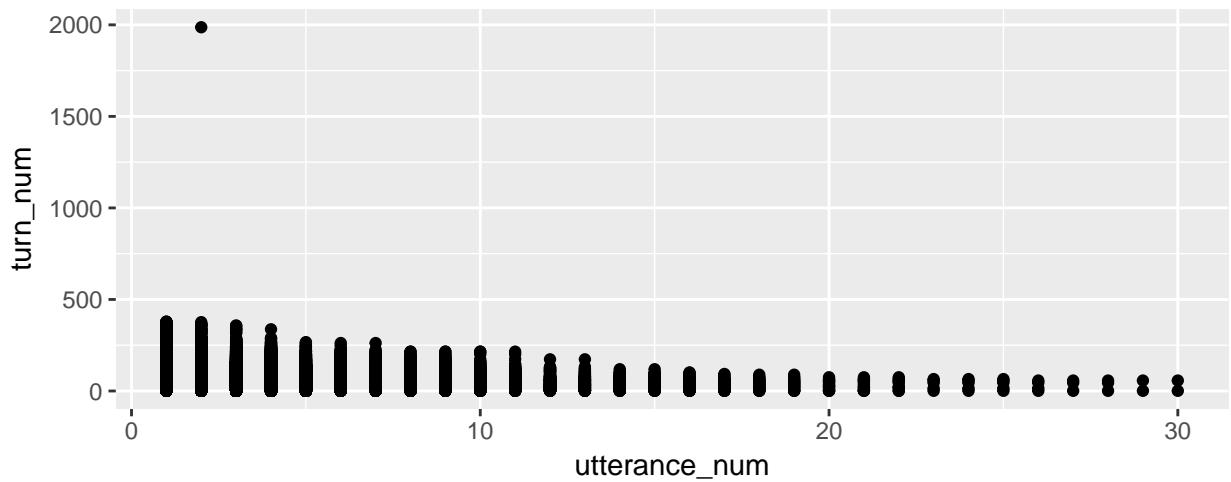
```
ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy))
```



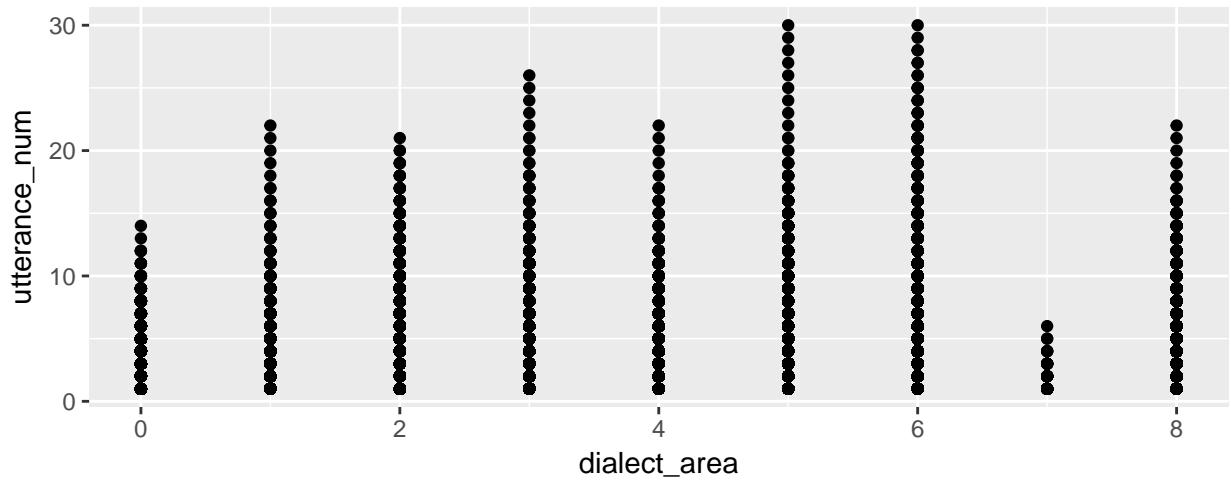
```
ggplot(data = df) + geom_point(mapping = aes(x = birth_year, y = turn_num))
```



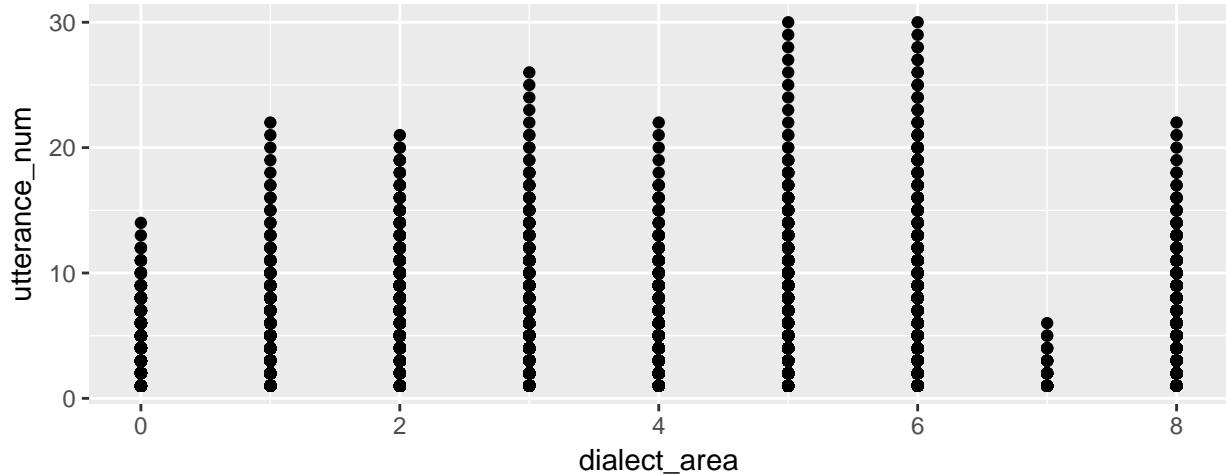
```
ggplot(data = df) + geom_point(mapping = aes(x = utterance_num, y = turn_num))
```



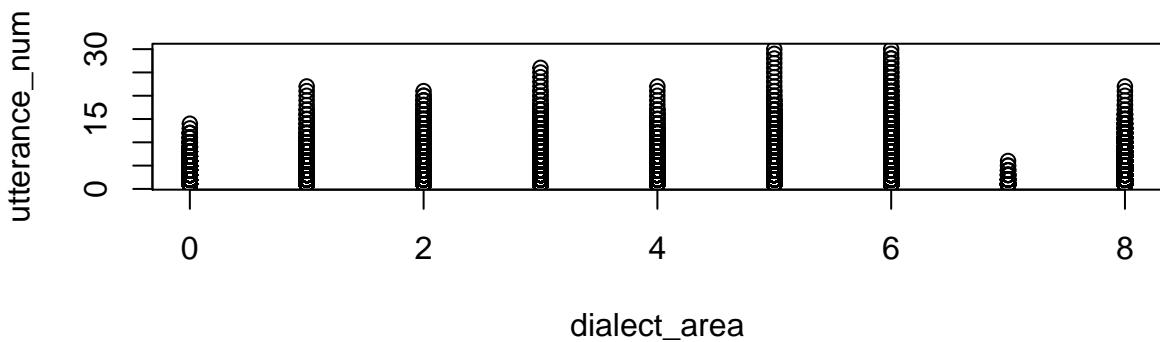
```
ggplot(data = df) + geom_point(mapping = aes(x = dialect_area, y = utterance_num))
```



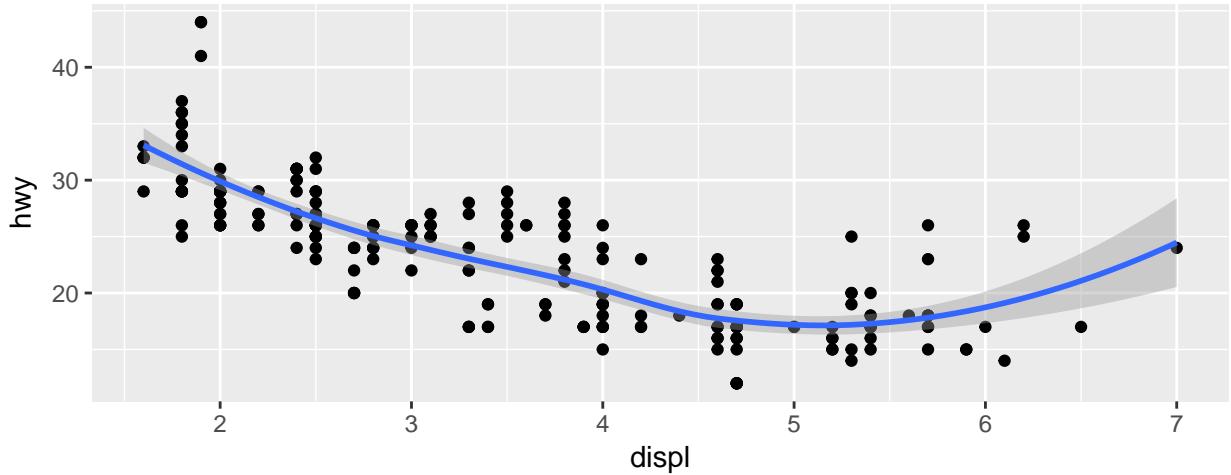
```
ggplot() + geom_point(data = df, mapping = aes(x = dialect_area, y = utterance_num))
```



```
plot(utterance_num ~ dialect_area, data = df)
```



```
ggplot() +
  geom_point(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_smooth(data = mpg, mapping = aes(x = displ, y = hwy))
## `geom_smooth()` using method = 'loess'
```

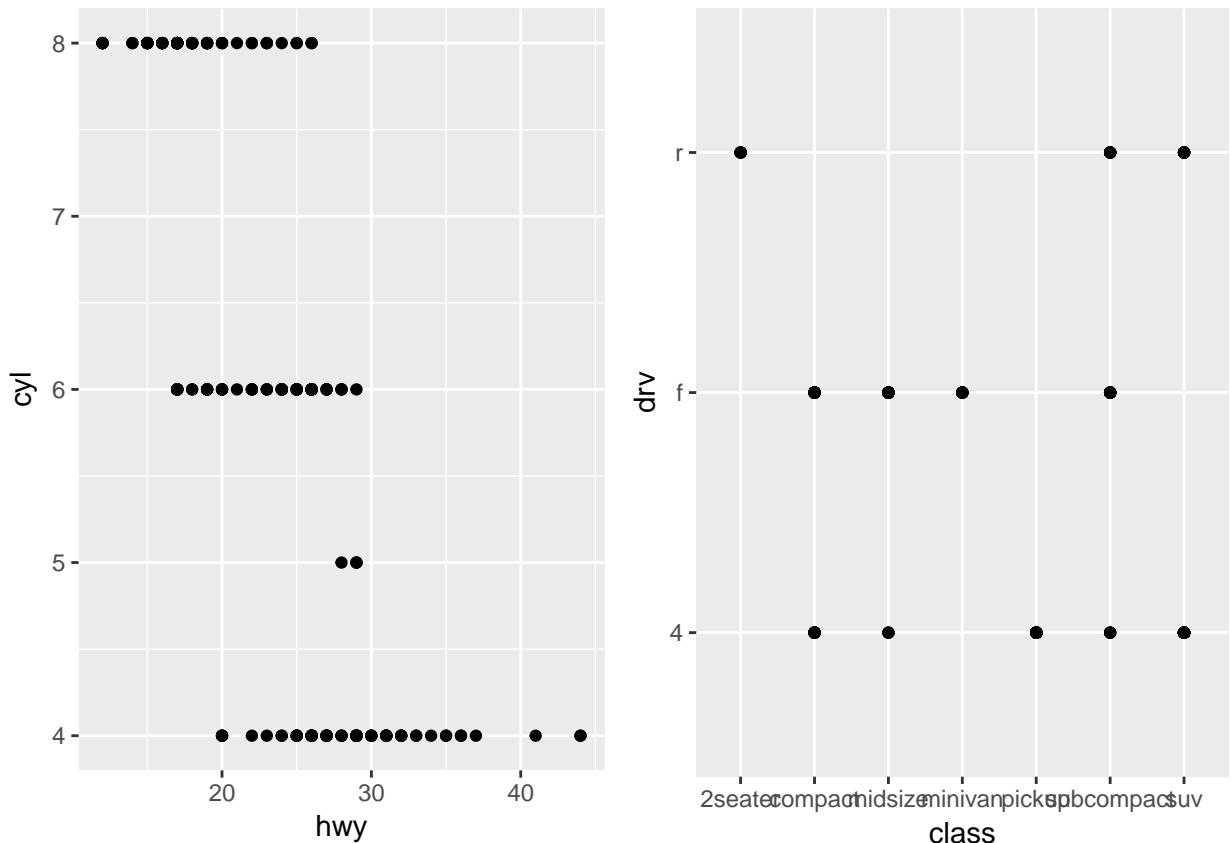


Exercises

```
dim(mpg) # 234 rows, 11 columns
```

drv: f = front-wheel drive, r = rear wheel drive, 4 = 4wd

```
p1 = ggplot(data = mpg) + geom_point(mapping = aes(x = hwy, y = cyl))
p2 = ggplot(data = mpg) + geom_point(mapping = aes(x = class, y = drv))
grid.arrange(p1, p2, ncol = 2)
```



Not useful because we can't see any patterns due to few data points?

3.3: Aesthetic Mappings

```
p1 = ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = class))

p2 = ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, size = class))

p3 = ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, alpha = class))

p4 = ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, shape = class))

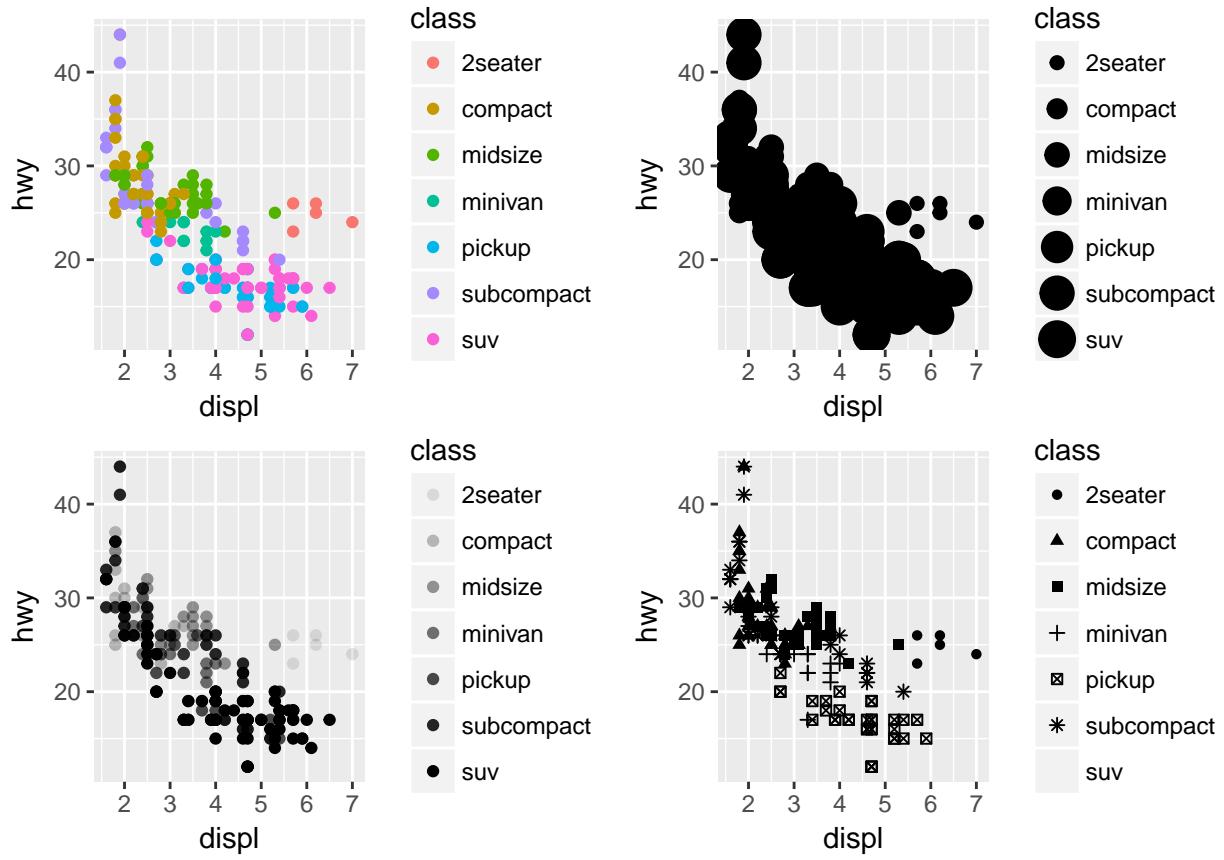
p5 = ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), color = "blue")

pushViewport(viewport(layout = grid.layout(2, 2)))
print(p1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(p2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))

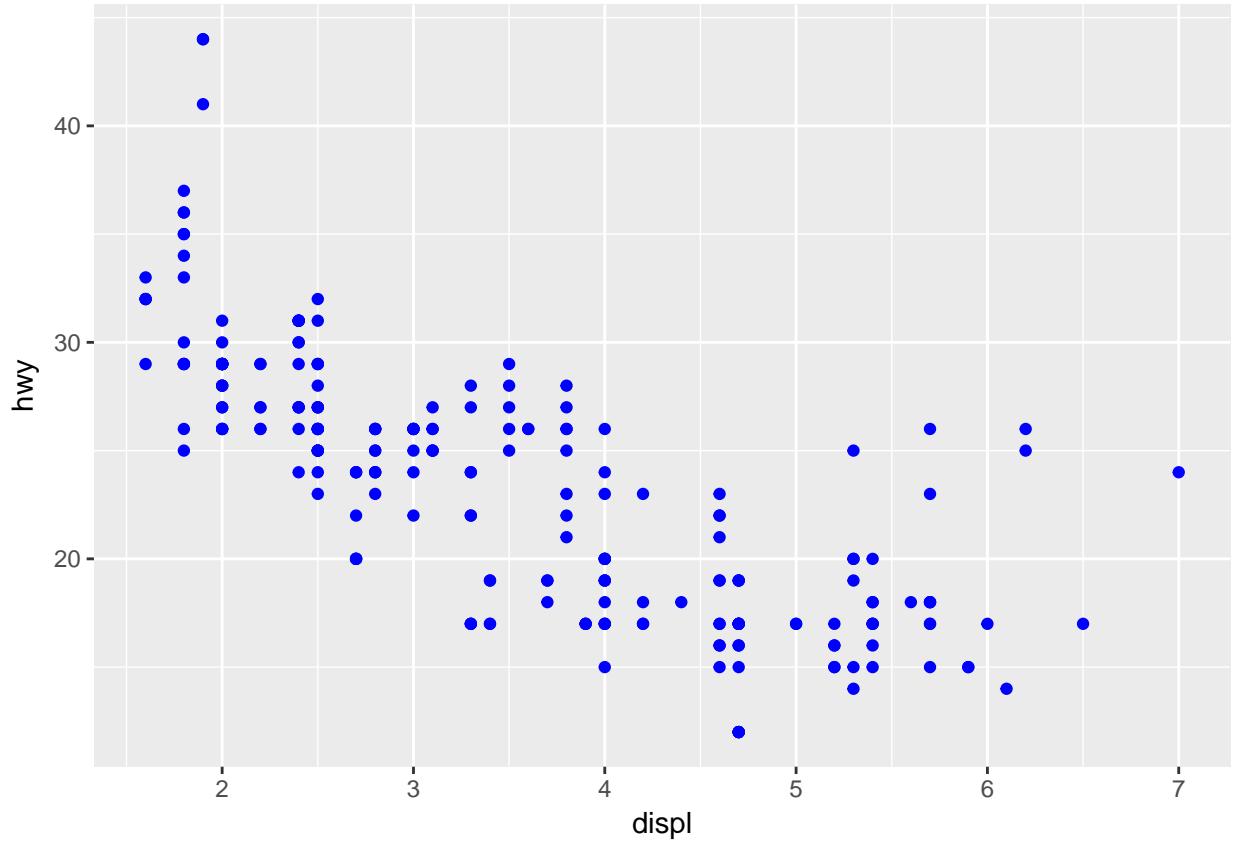
## Warning: Using size for a discrete variable is not advised.
print(p3, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
print(p4, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))

## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have 7.
## Consider specifying shapes manually if you must have them.

## Warning: Removed 62 rows containing missing values (geom_point).
```



```
show(p5)
```



Exercises

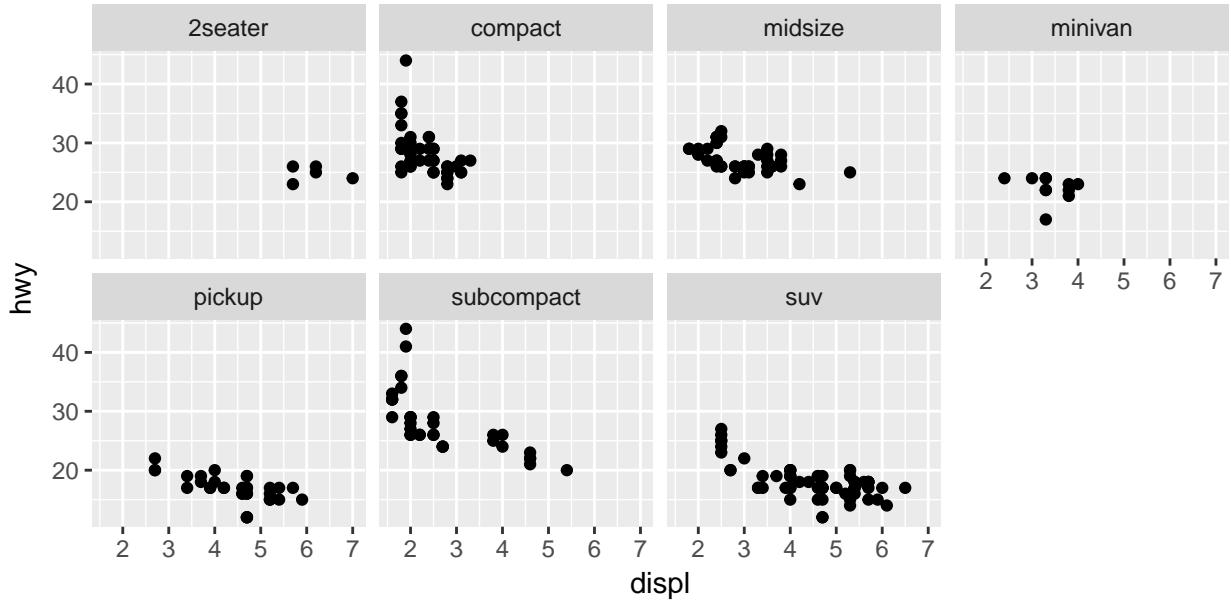
1. The correct code is `ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy), color = "blue")`.
2. categorical vs. continuous
- 3.
- 4.
5. stroke: to modify the width of the border of a shape with a border
- 6.

3.4: Common problems

3.5: Facets

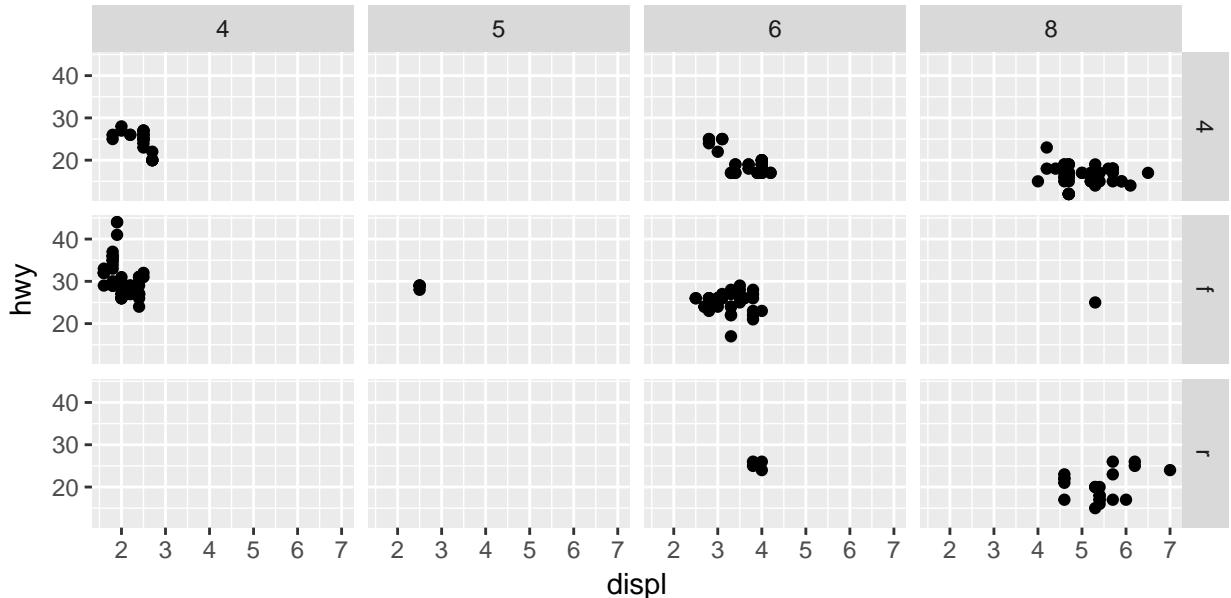
A facet is a subplot that displays one subset of the data.

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```



`facet_grid()` to facet my plot on the combo of two variables.

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



```
# facet_grid(. ~ cyl) if you prefer to not facet in the rows of columns direction
```

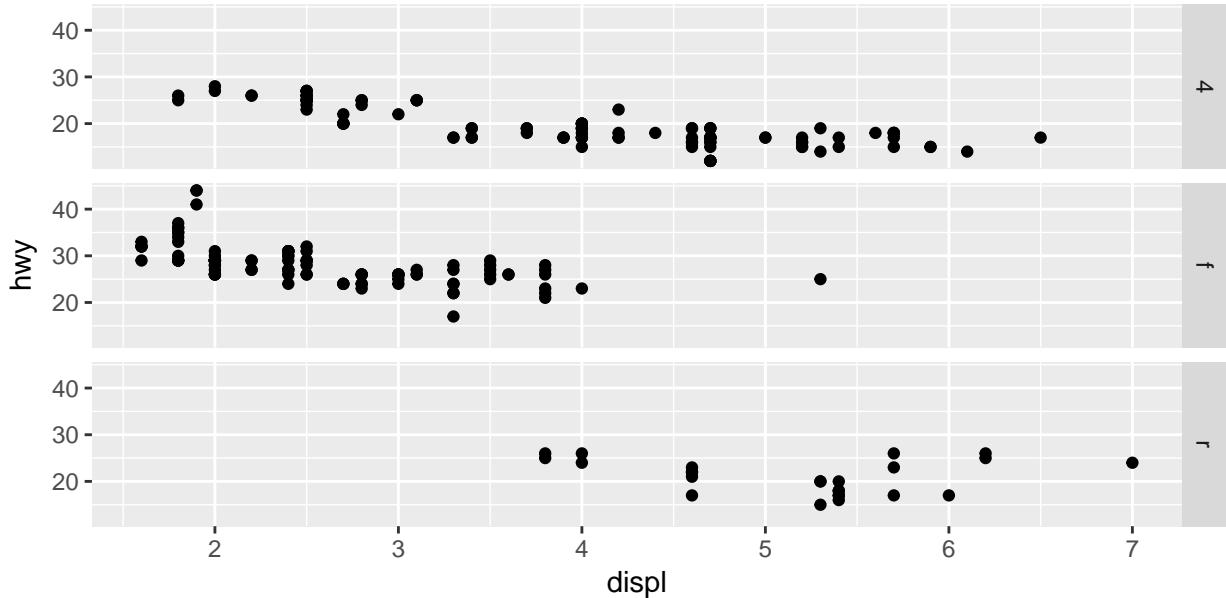
Exercises

1.

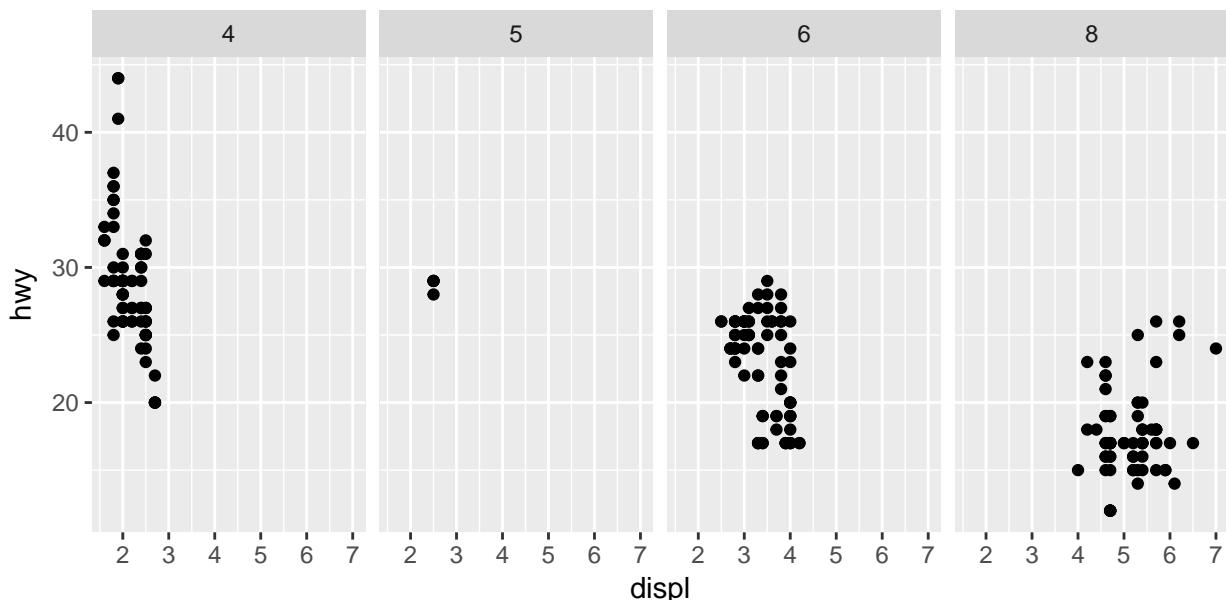
2.

3.

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

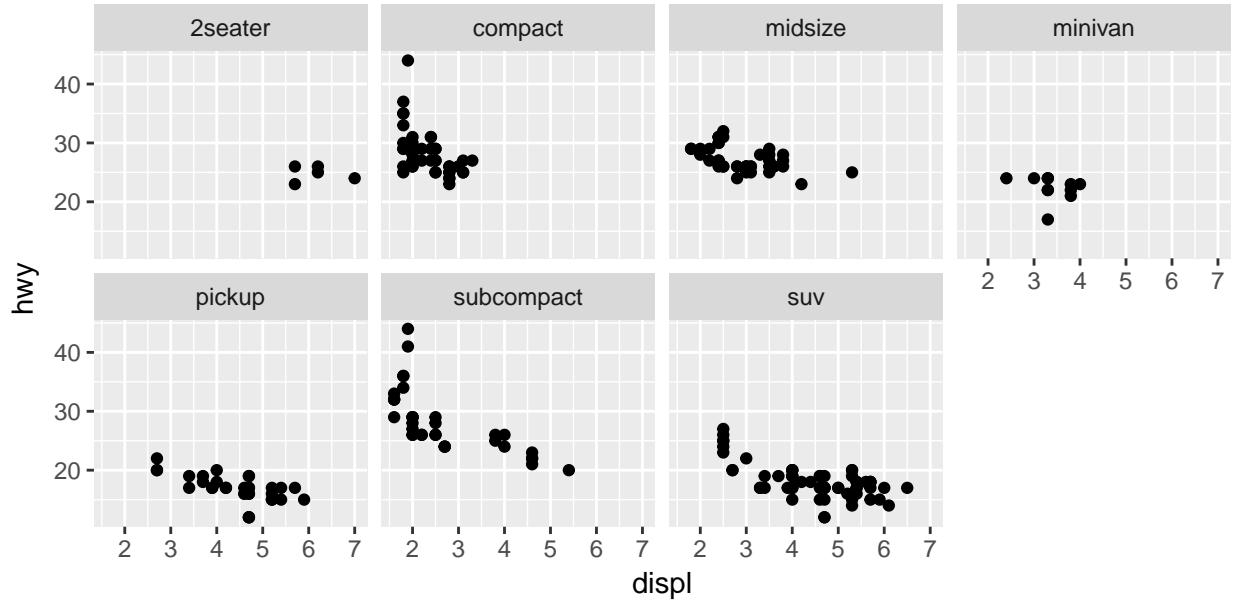


```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_grid(. ~ cyl)
```



4.

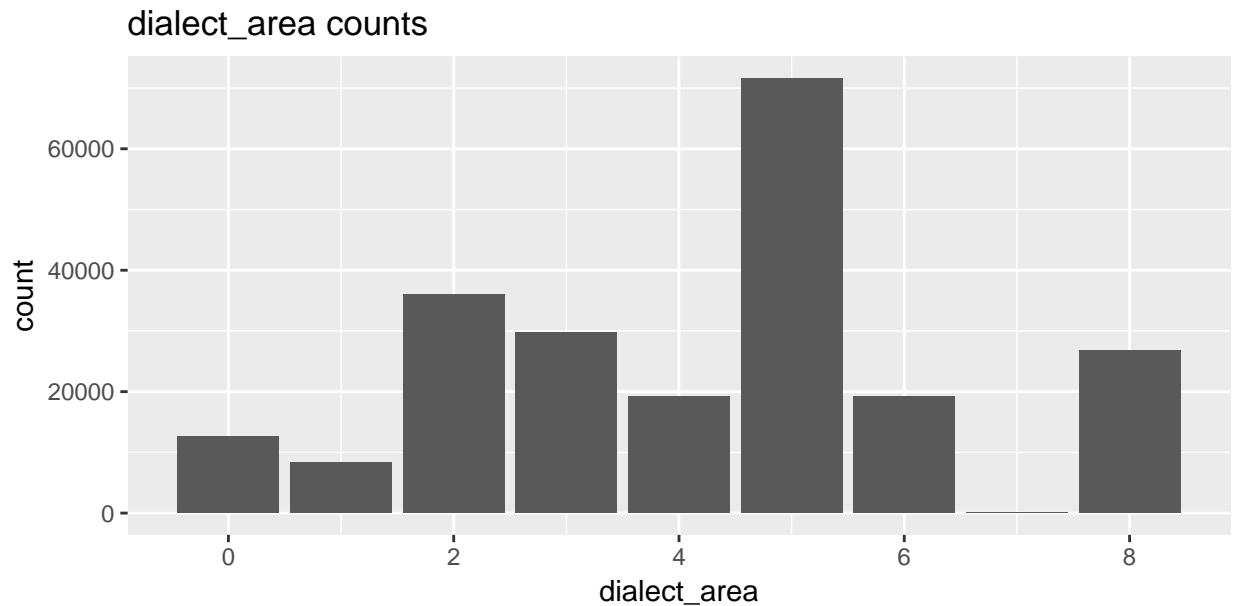
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_wrap(~ class, nrow = 2)
```



3.6: Geometric objects

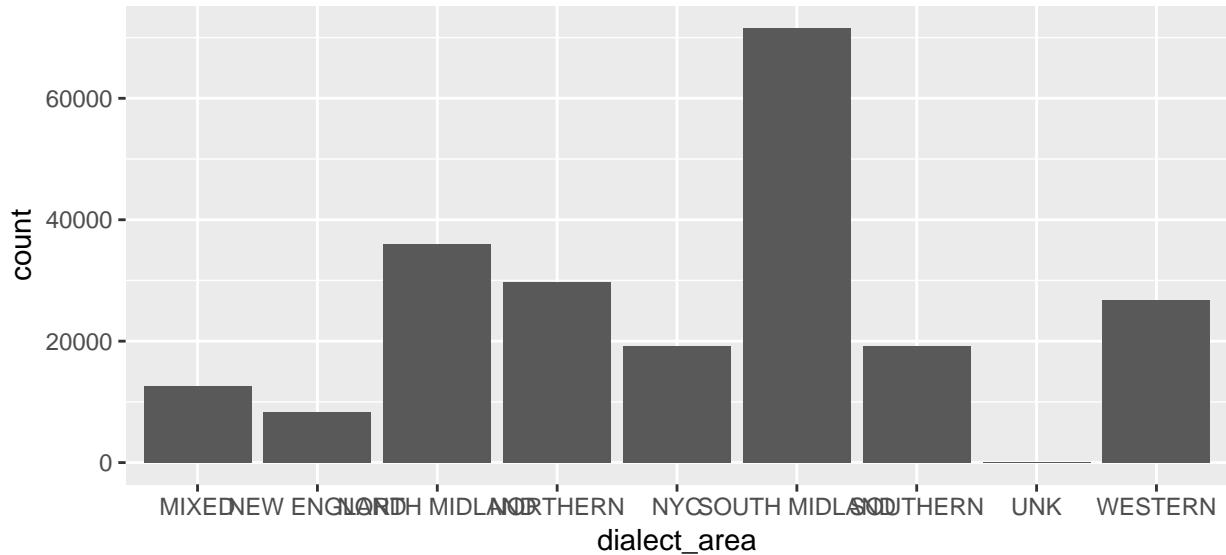
3.7: Statistical transformations

```
ggplot(data = df) + geom_bar(mapping = aes(x = dialect_area)) + labs(title = 'dialect_area counts')
```



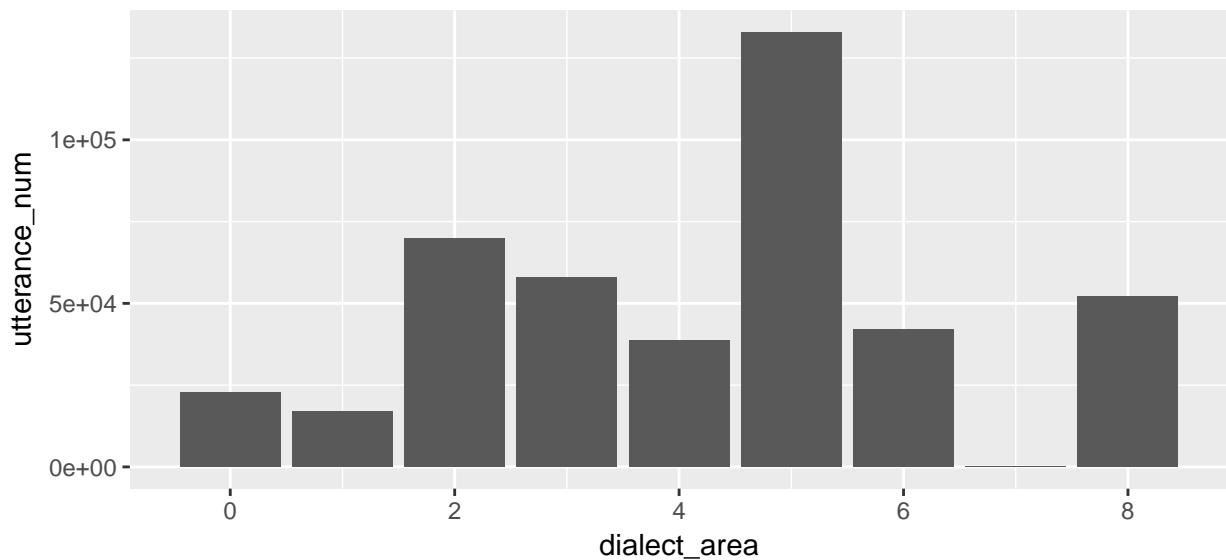
```
ggplot(data = swda) + geom_bar(mapping = aes(x = dialect_area)) + labs(title = 'dialect_area counts')
```

dialect_area counts

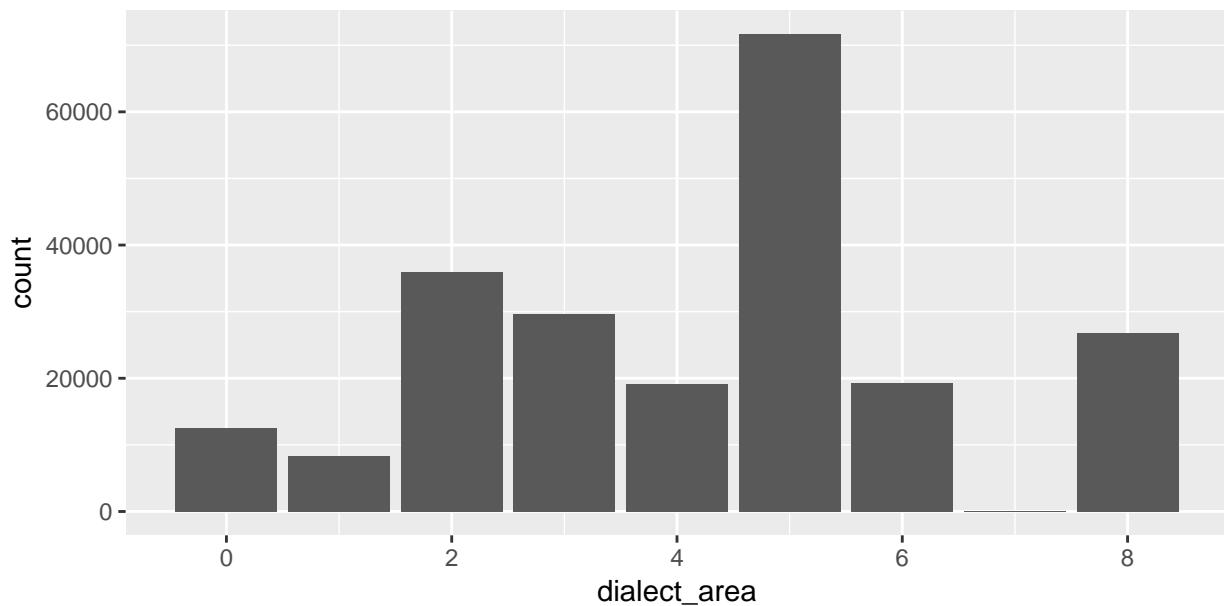


```
ggplot(data = df) + geom_bar(mapping = aes(x = dialect_area, y = utterance_num), stat = "identity") + 1
```

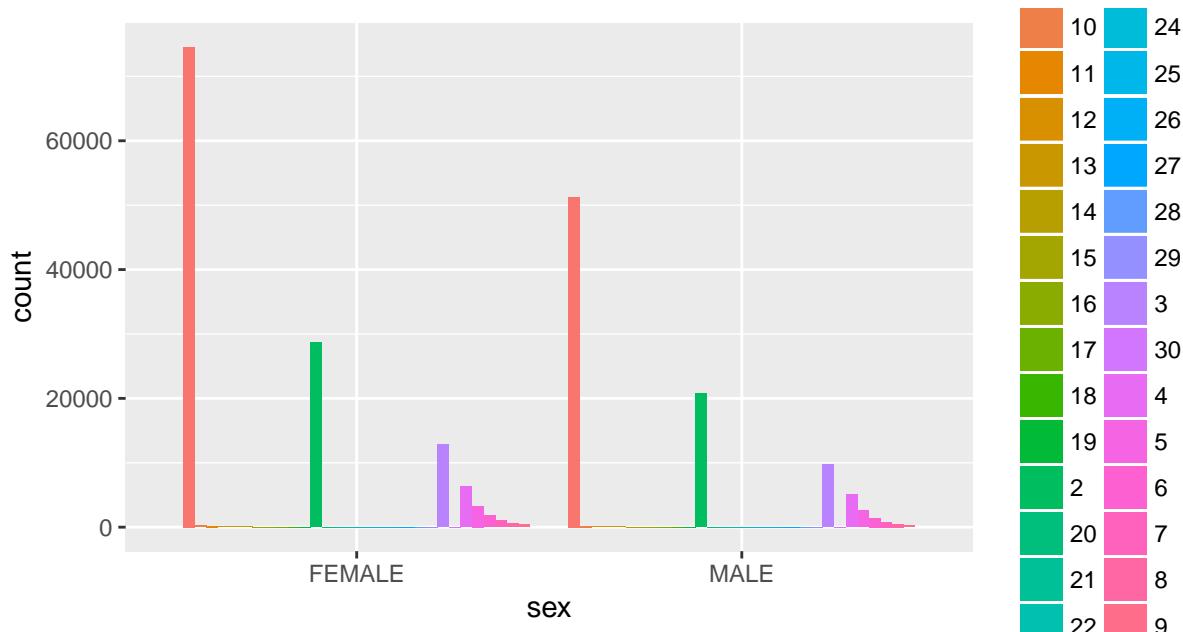
dialect_area ~ utterance_count



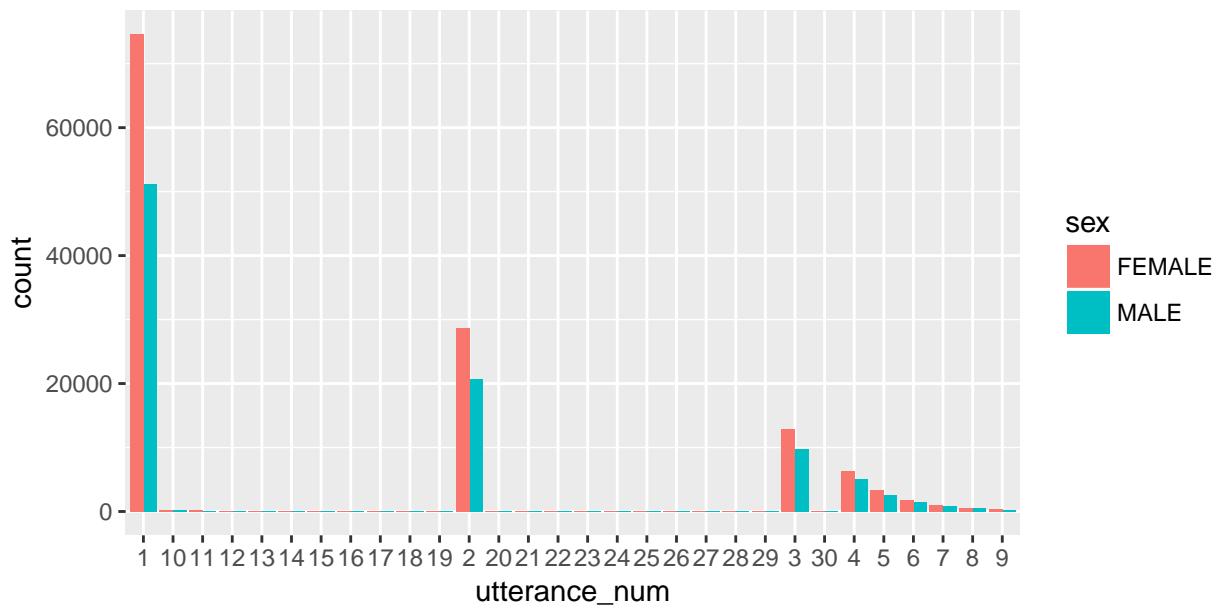
```
ggplot(df, aes(dialect_area, ..count..)) + geom_bar(aes(fill = utterance_num), position = "dodge")
```



```
ggplot(swda, aes(sex, ..count..)) + geom_bar(aes(fill = utterance_num), position = "dodge")
```



```
ggplot(swda, aes(utterance_num, ..count..)) + geom_bar(aes(fill = sex), position = "dodge")
```



3.8: Position adjustments

3.9: Coordinate systems

3.10: