



DATA VISUALIZATION WITH GGPLOT2

Bar Plots

Chapter Content

- Common pitfalls
- Best way to represent data

Bar plot

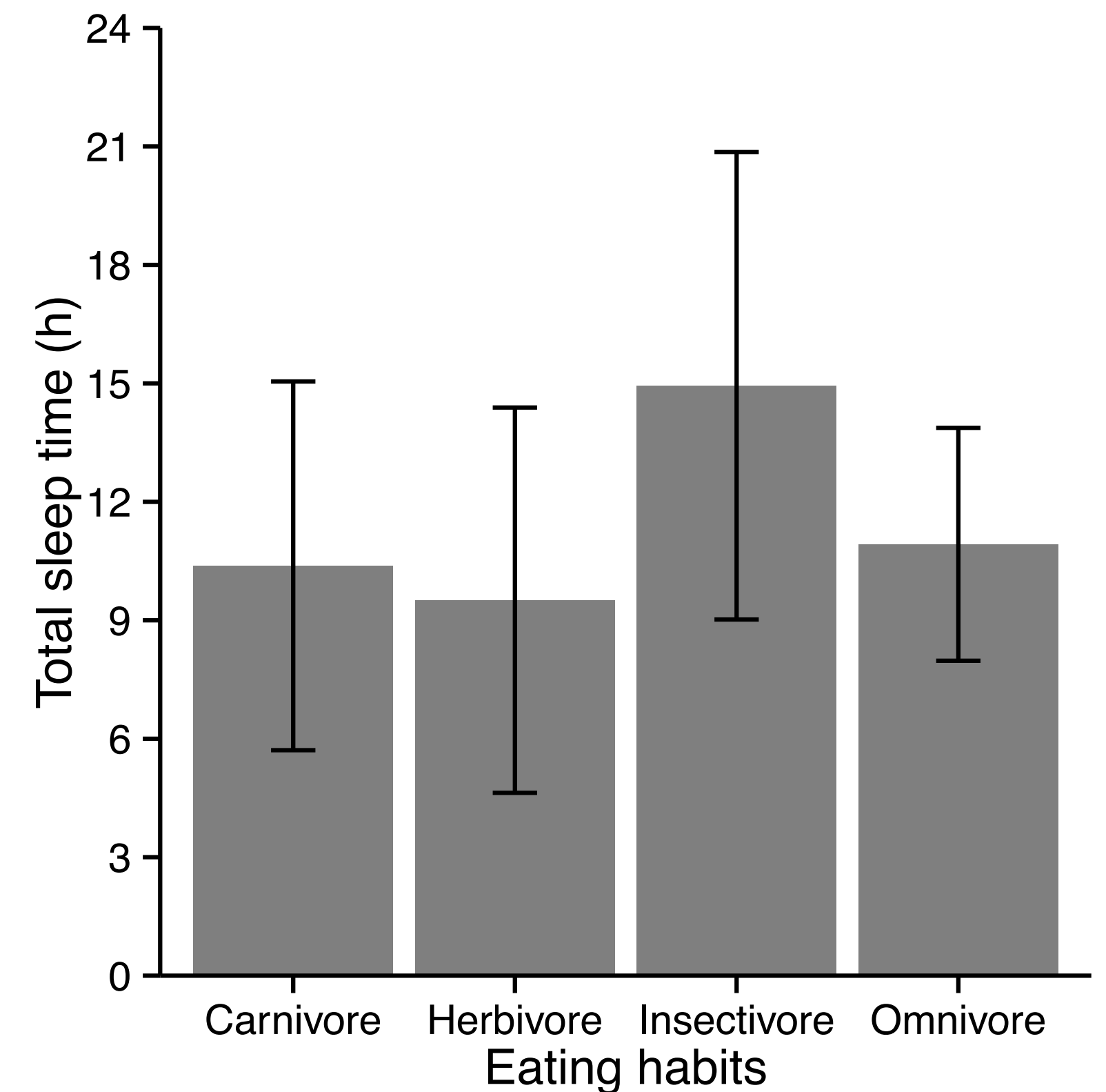
- Two types
 - Absolute values
 - Distribution

Mammalian sleep

```
> str(sleep)
'data.frame': 76 obs. of  3 variables:
 $vore : Factor w/ 4 levels "Carnivore","Herbivore",...: 1 4 2 ...
 $total: num  12.1 17 14.4 14.9 4 14.4 8.7 10.1 3 5.3 ...
 $rem  : num  NA 1.8 2.4 2.3 0.7 2.2 1.4 2.9 NA 0.6 ...
```

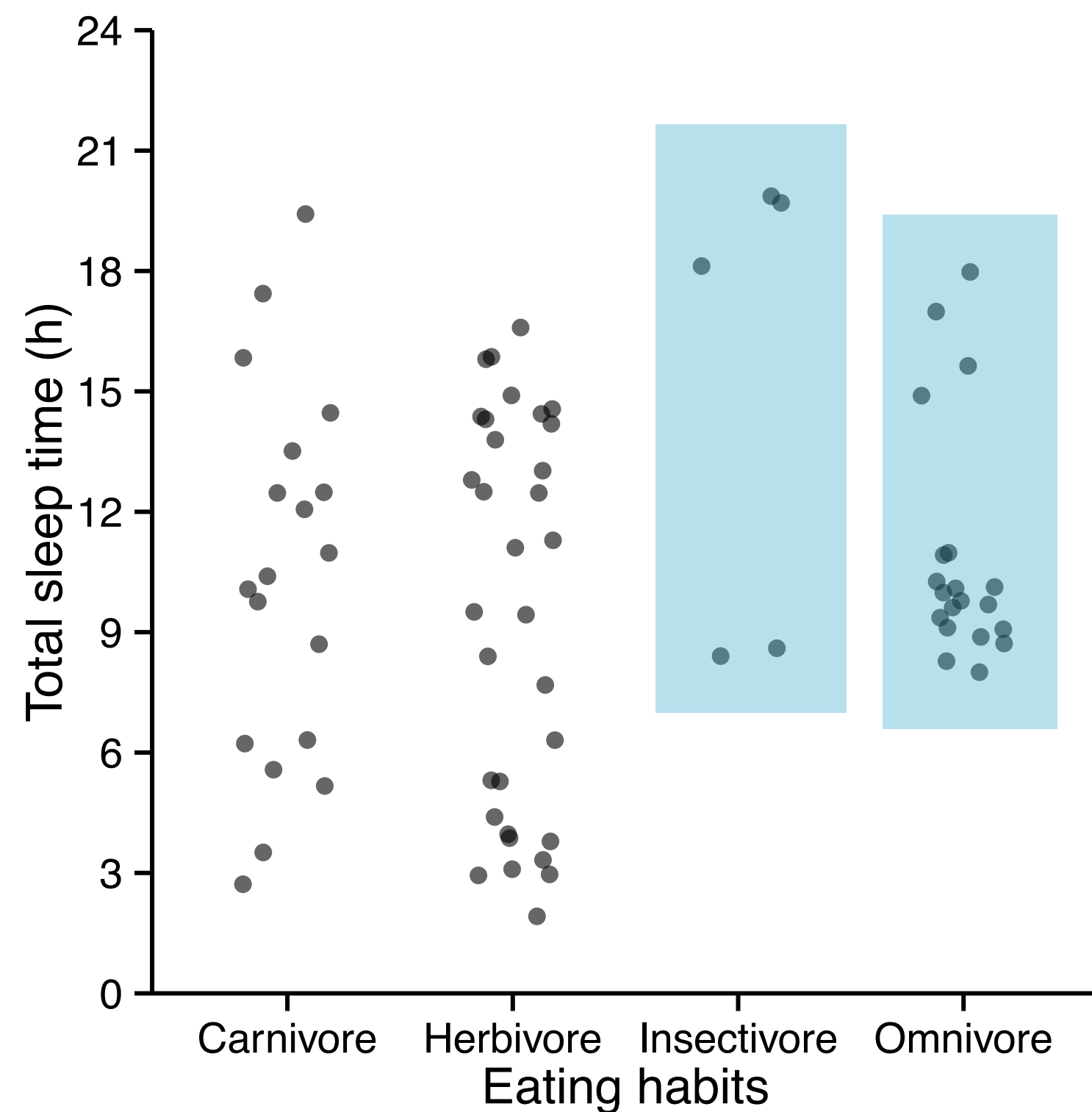
Dynamite plot

```
> d <- ggplot(sleep, aes(vore, total)) +  
  scale_y_continuous("Total sleep time (h)",  
    limits = c(0, 24),  
    breaks = seq(0, 24, 3),  
    expand = c(0, 0)) +  
  scale_x_discrete("Eating habits") +  
  theme_classic()  
  
> d +  
  stat_summary(fun.y = mean, geom = "bar",  
    fill = "grey50") +  
  stat_summary(fun.data = mean_sdl, mult = 1,  
    geom = "errorbar", width = 0.2)
```



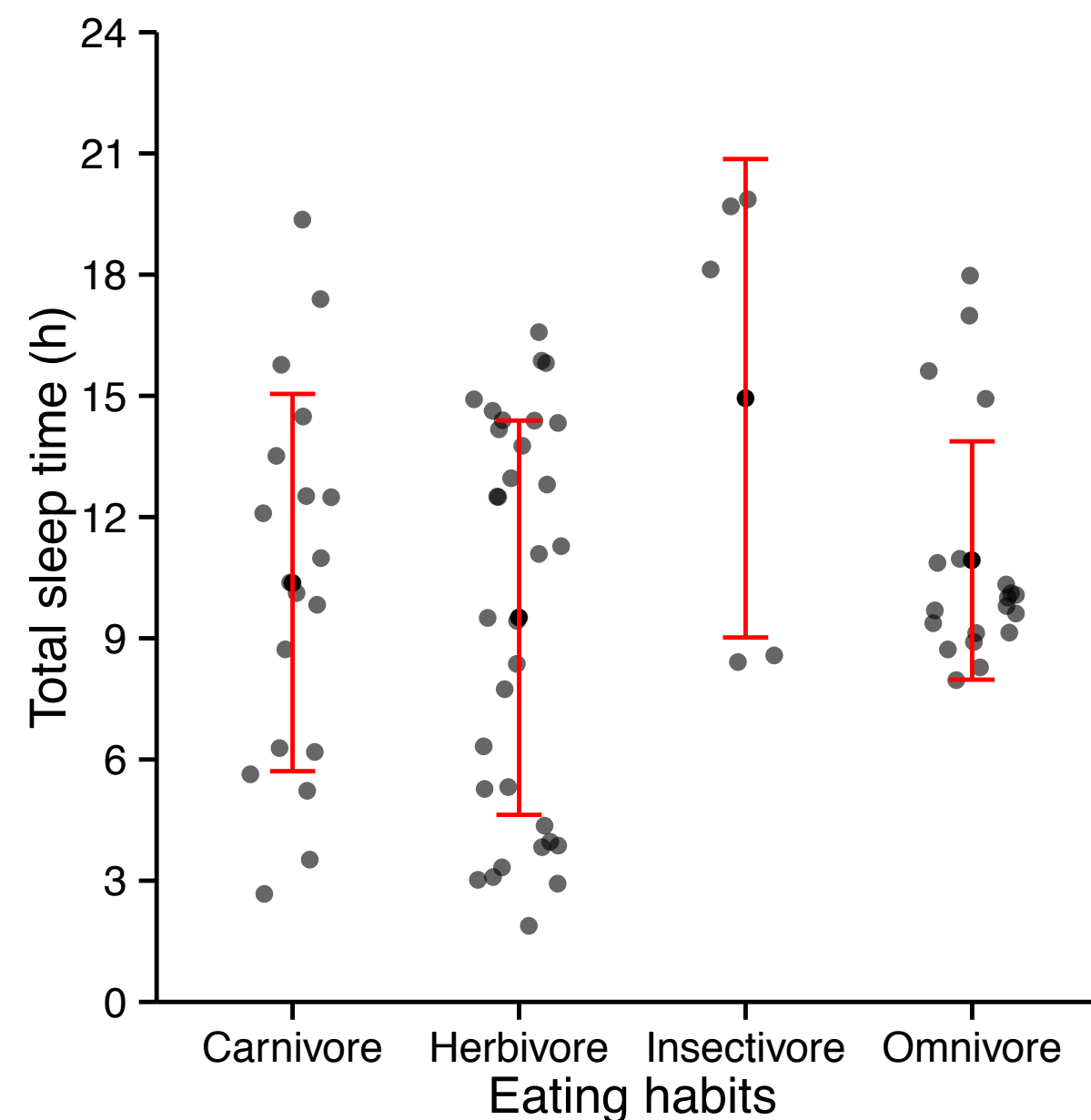
Individual data points

```
> d +  
  geom_point(alpha = 0.6, position = position_jitter(width = 0.2))
```



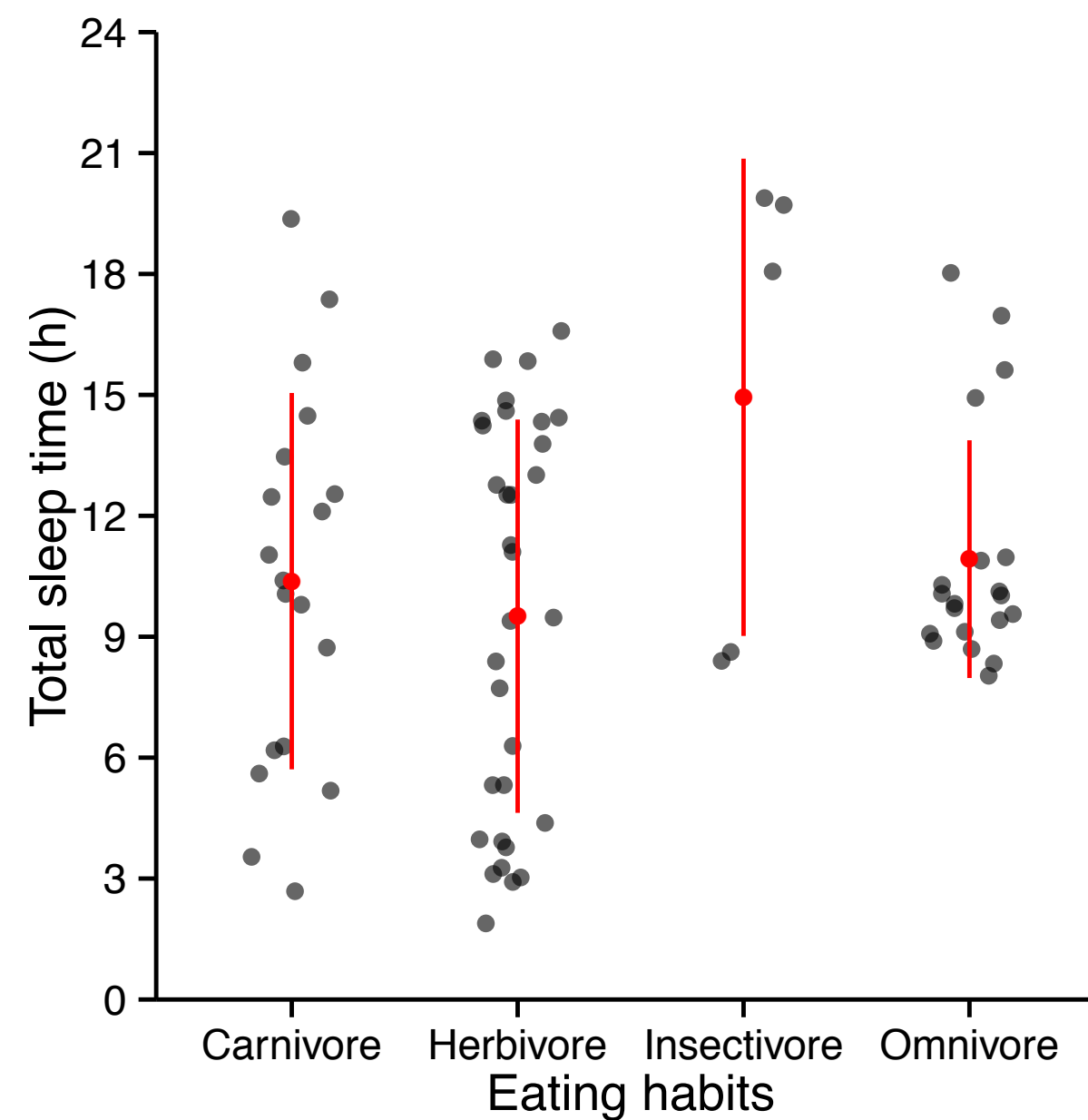
errorbar

```
> d +  
  geom_point(alpha = 0.6, position = position_jitter(width = 0.2)) +  
  stat_summary(fun.y = mean, geom = "point", fill = "red") +  
  stat_summary(fun.data = mean_sdl, mult = 1, geom = "errorbar",  
              width = 0.2, col = "red")
```



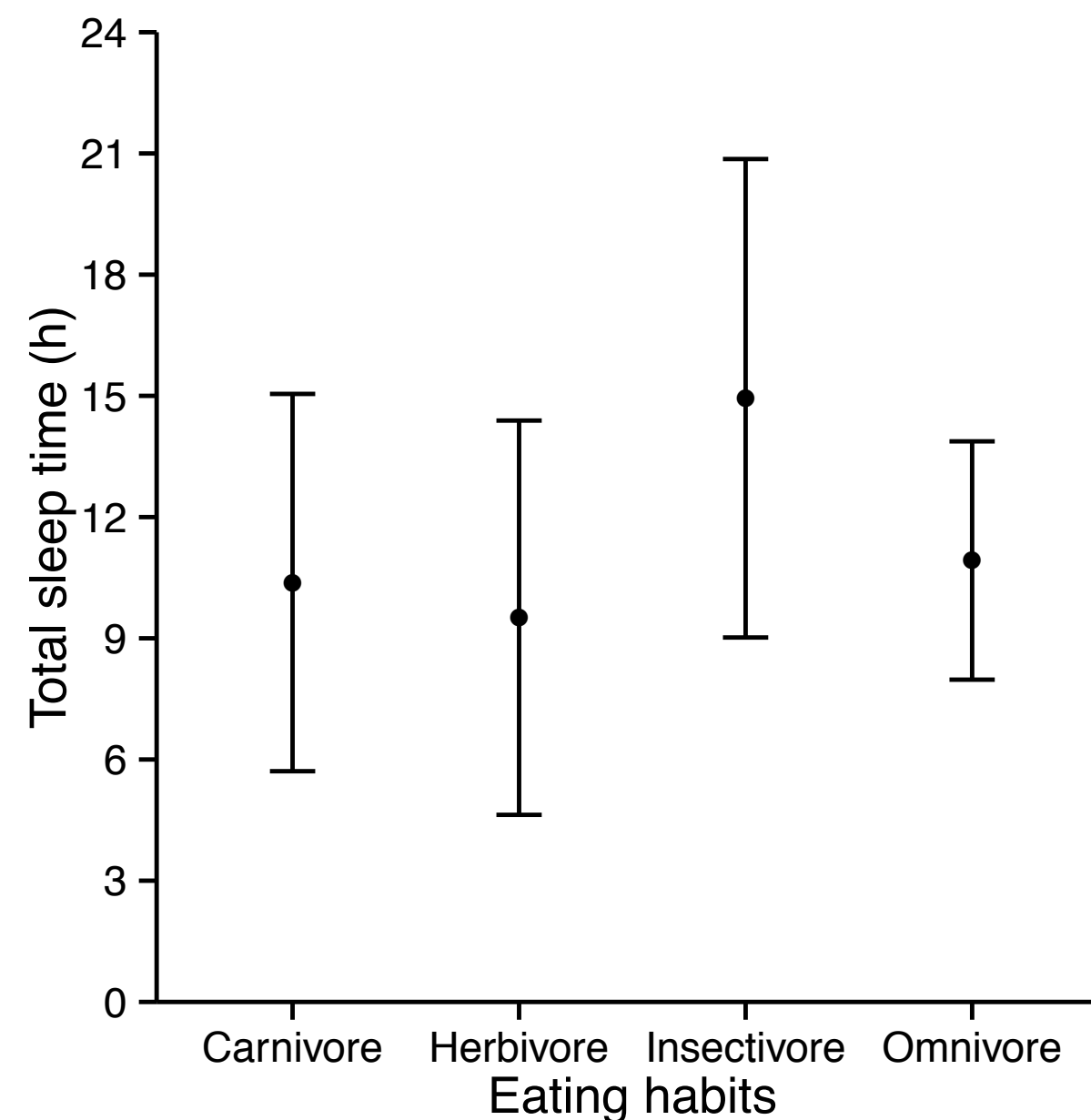
pointrange

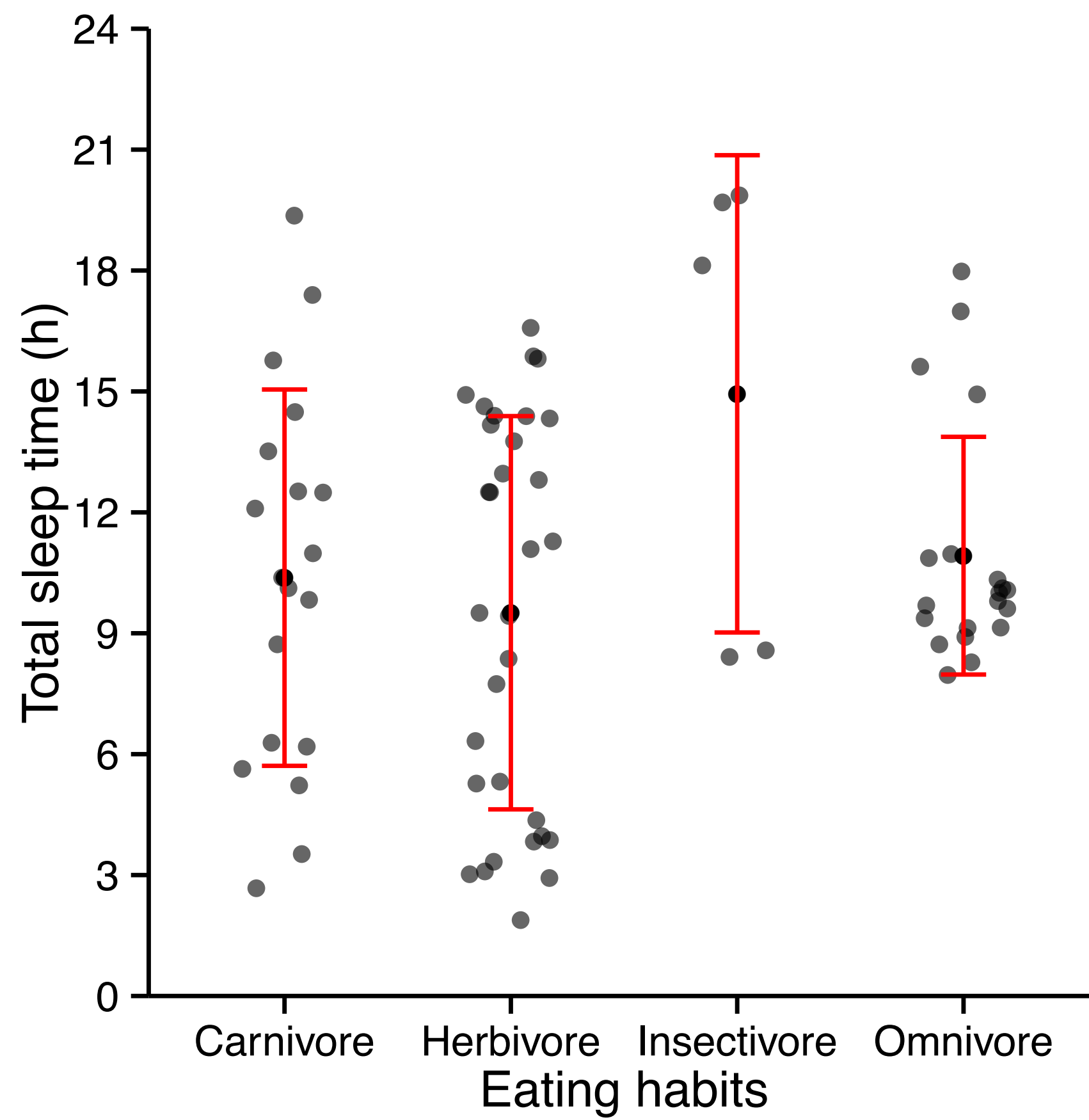
```
> d +  
  geom_point(alpha = 0.6, position = position_jitter(width = 0.2)) +  
  stat_summary(fun.data = mean_sdl, mult = 1, width = 0.2, col = "red")
```



Without data points

```
> d +  
  stat_summary(fun.y = mean, geom = "point") +  
  stat_summary(fun.data = mean_sdl, mult = 1,  
              geom = "errorbar", width = 0.2)
```







DATA VISUALIZATION WITH GGLOT2

Let's practice!

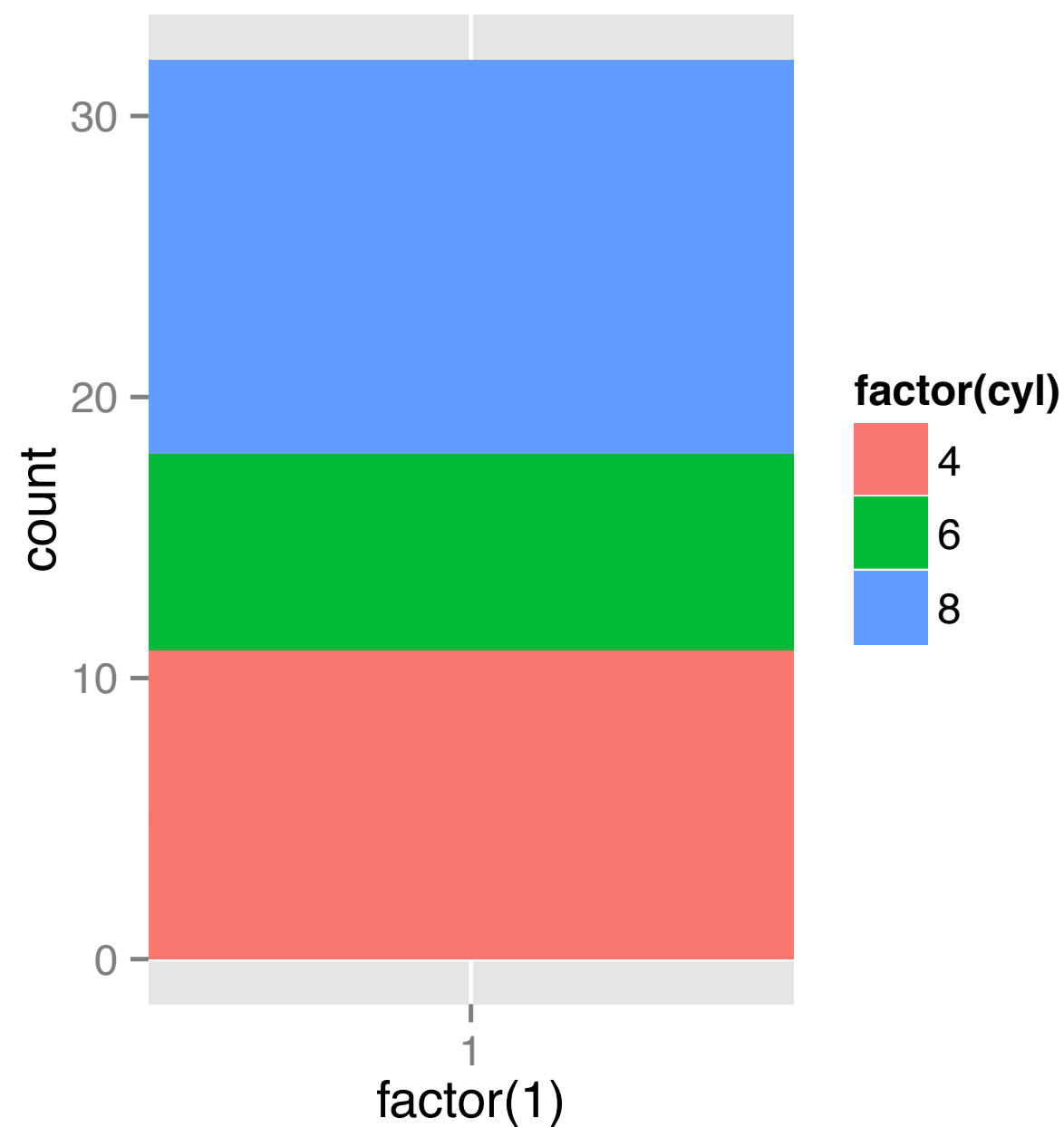


DATA VISUALIZATION WITH GGLOT2

Pie Charts

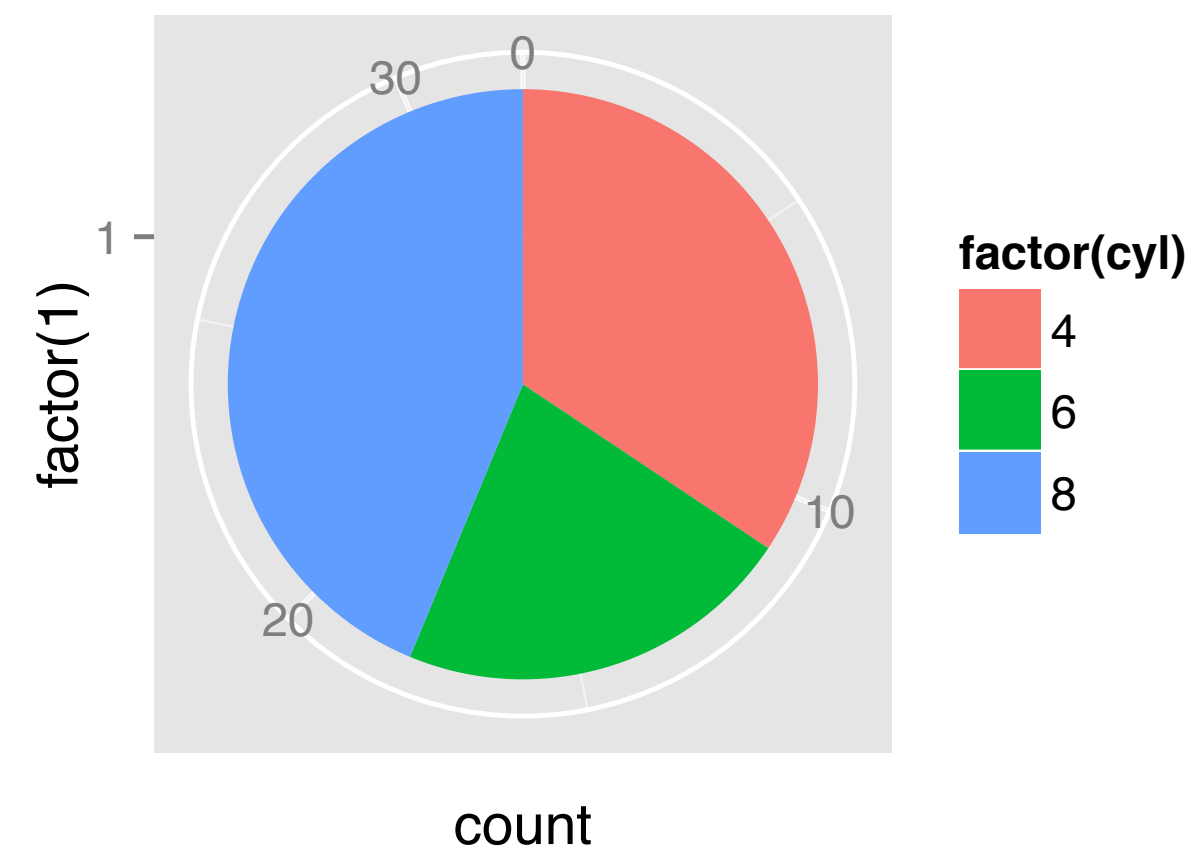
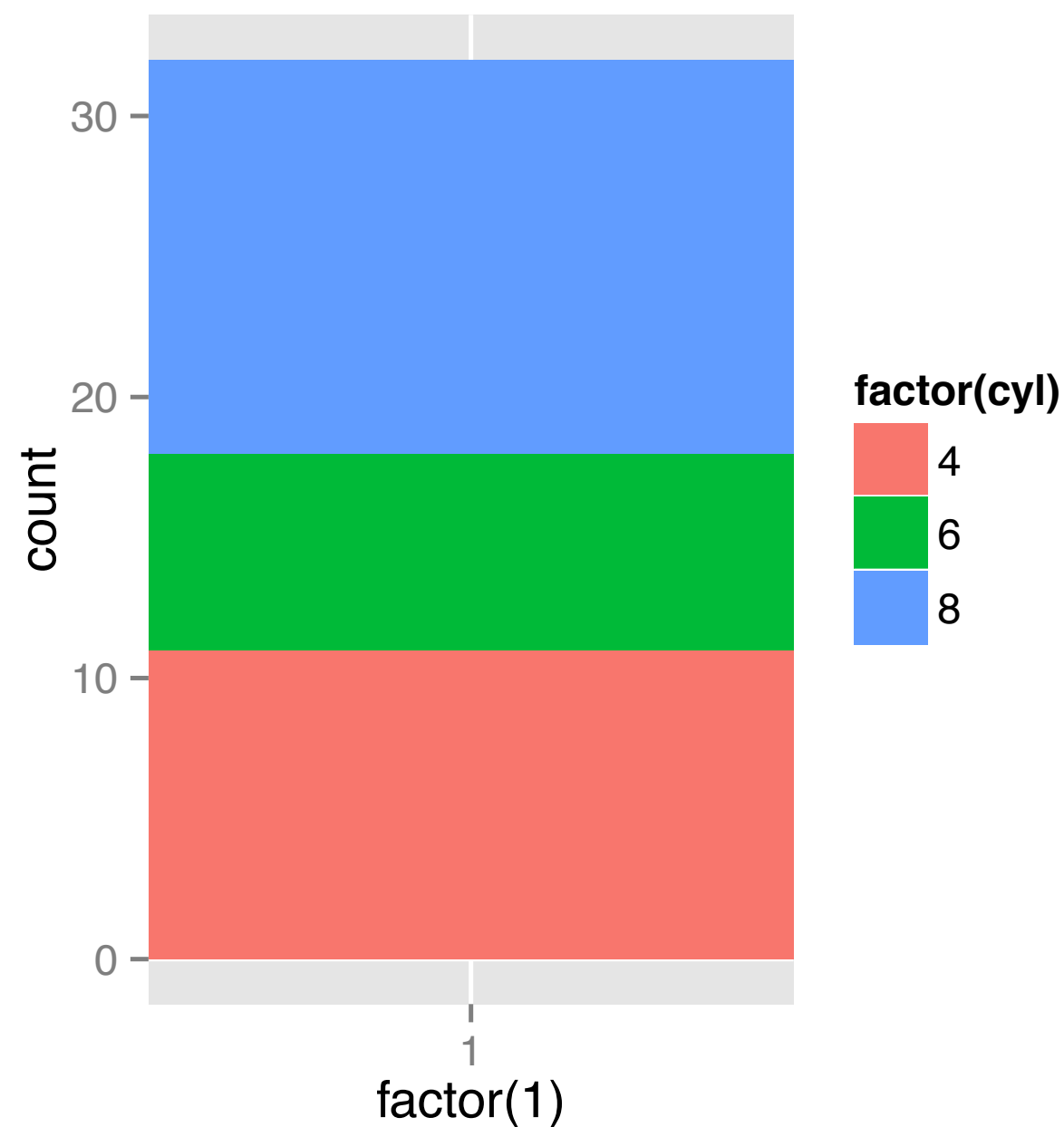
Stacked bar chart ...

```
> ggplot(mtcars, aes(x = factor(1), fill = factor(cyl))) +  
  geom_bar(width = 1)
```

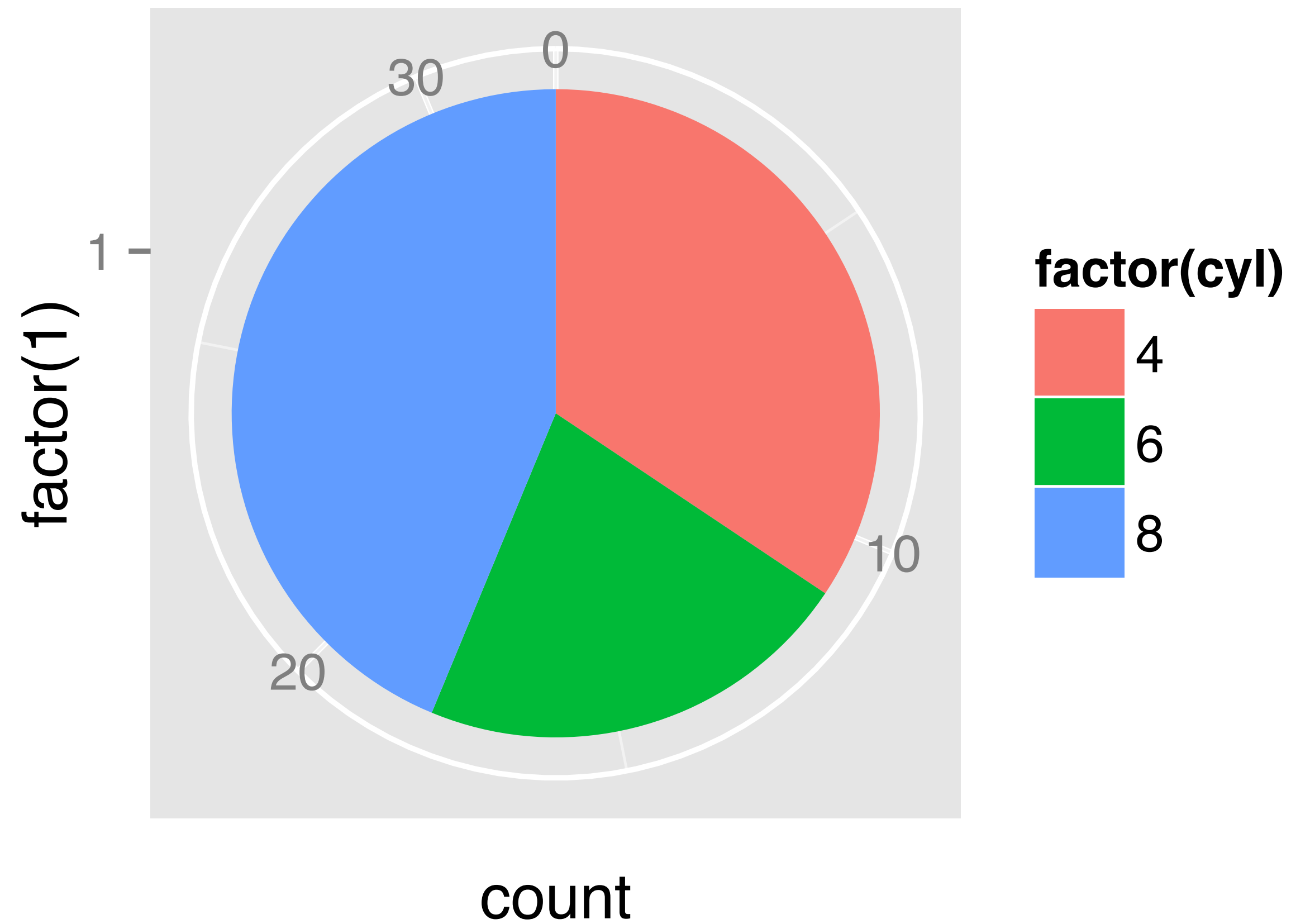


... pie chart

```
> ggplot(mtcars, aes(x = factor(1), fill = factor(cyl))) +  
  geom_bar(width = 1)  
> ggplot(mtcars, aes(x = factor(1), fill = factor(cyl))) +  
  geom_bar(width = 1) +  
  coord_polar(theta = "y")
```



Parts-of-a-whole

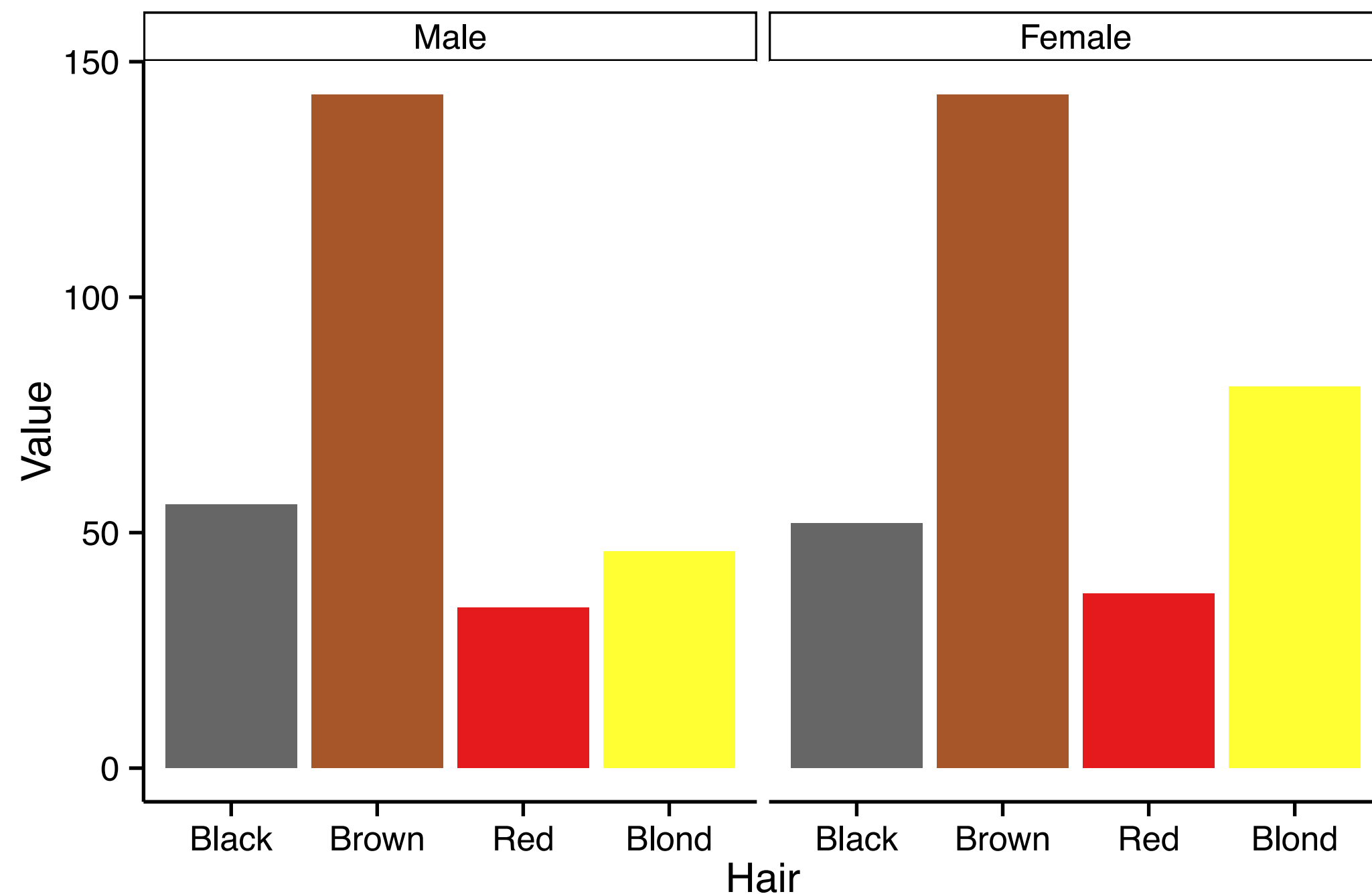


HairCol

```
> HairCol
  Hair    Sex Value fillin    n    nprop
1 Black  Male   56 #666666 279 0.4712838
2 Brown  Male  143 #A65628 279 0.4712838
3  Red   Male   34 #E41A1C 279 0.4712838
4 Blond  Male   46 #FFFF33 279 0.4712838
5 Black Female   52 #666666 313 0.5287162
6 Brown Female  143 #A65628 313 0.5287162
7  Red  Female   37 #E41A1C 313 0.5287162
8 Blond Female   81 #FFFF33 313 0.5287162
```


HairCol - Bar Charts

```
> ggplot(HairCol, aes(x = Hair, y = Value, fill = fillin)) +  
  geom_bar(stat = "identity", position = "dodge") +  
  facet_grid(. ~ Sex) +  
  scale_fill_identity() +  
  theme_classic()
```

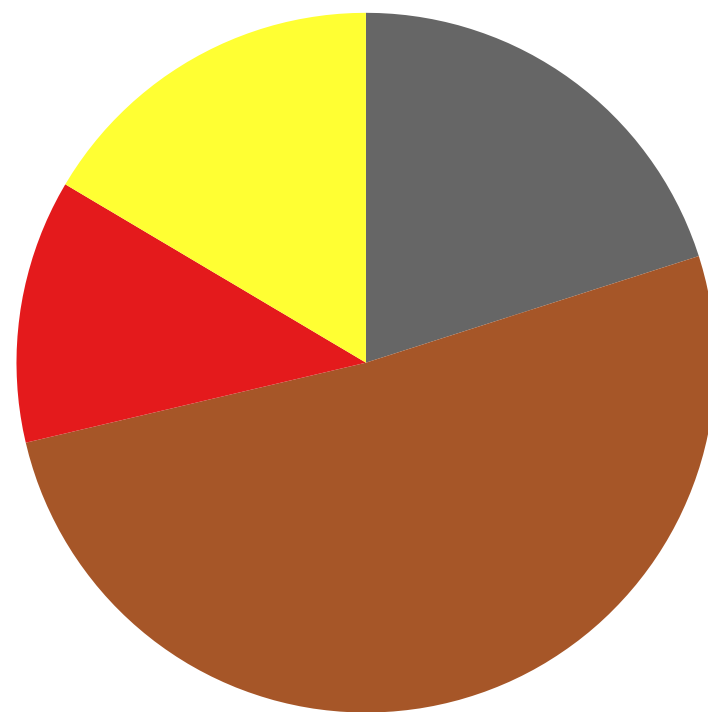


Hard to reveal interesting trends
Difference in total counts is unclear

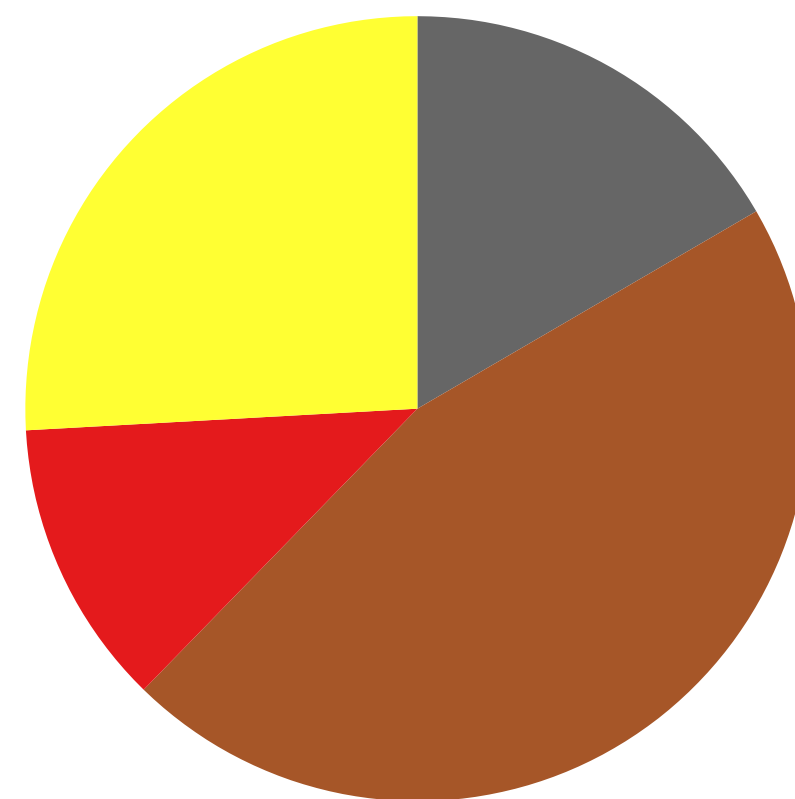
HairCol - Pie Charts

```
> ggplot(HairCol, aes(x = n/2, y = Value, fill = fillin, width = n)) +  
  geom_bar(stat = "identity", position = "fill") +  
  facet_grid(. ~ Sex) +  
  scale_fill_identity() +  
  coord_polar(theta = "y") +  
  theme(...)
```

Male



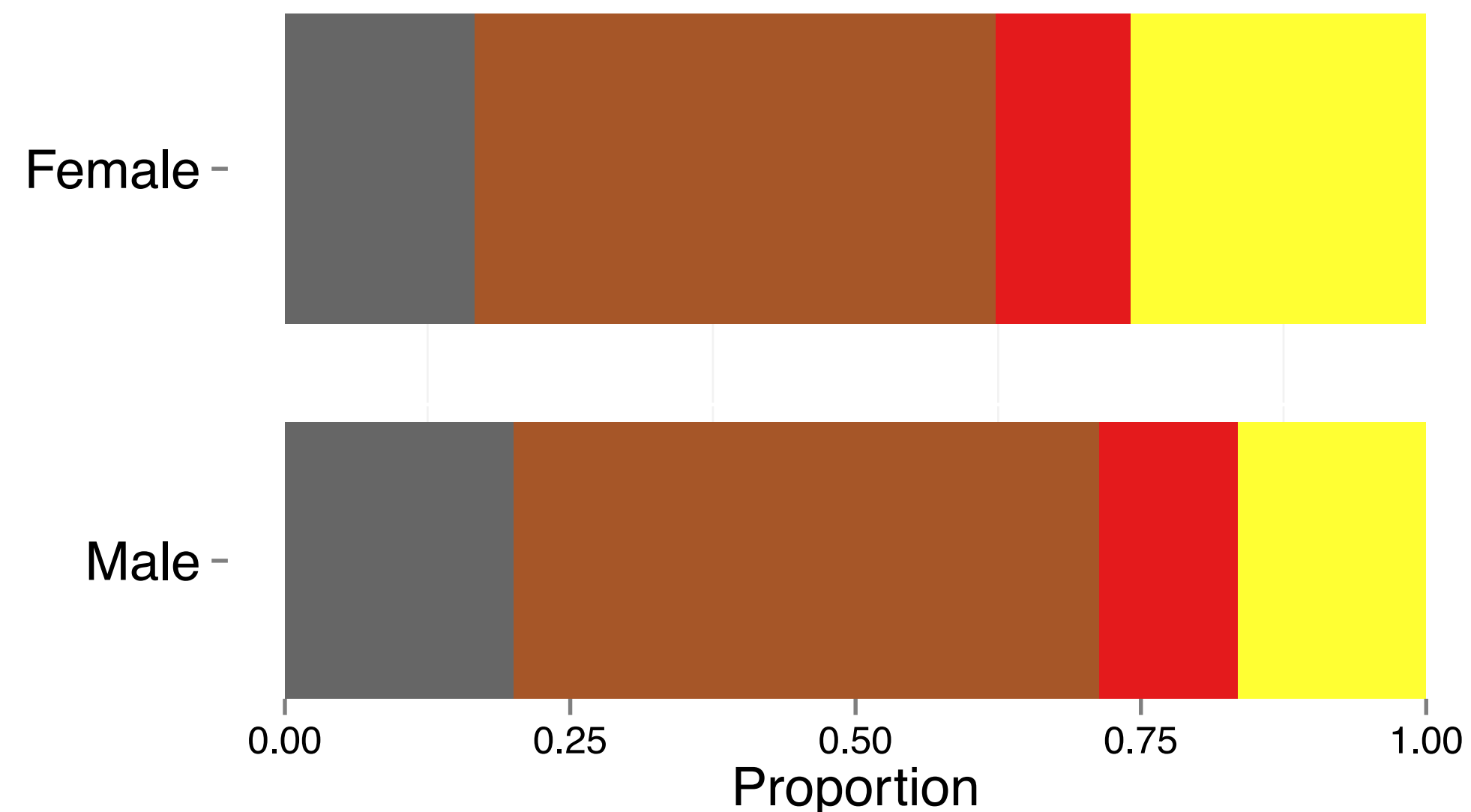
Female



angle, area, length
mediocre encoding elements

Alternative

```
> ggplot(HairCol, aes(x = Sex, y = Value, fill = fillin, width = nprop)) +  
  geom_bar(stat = "identity", position = "fill") +  
  scale_y_continuous("Proportion") +  
  scale_x_discrete("", expand = c(0, 0)) +  
  scale_fill_identity() +  
  coord_flip() +  
  theme(...)
```





DATA VISUALIZATION WITH GGPLOT2

Let's practice!



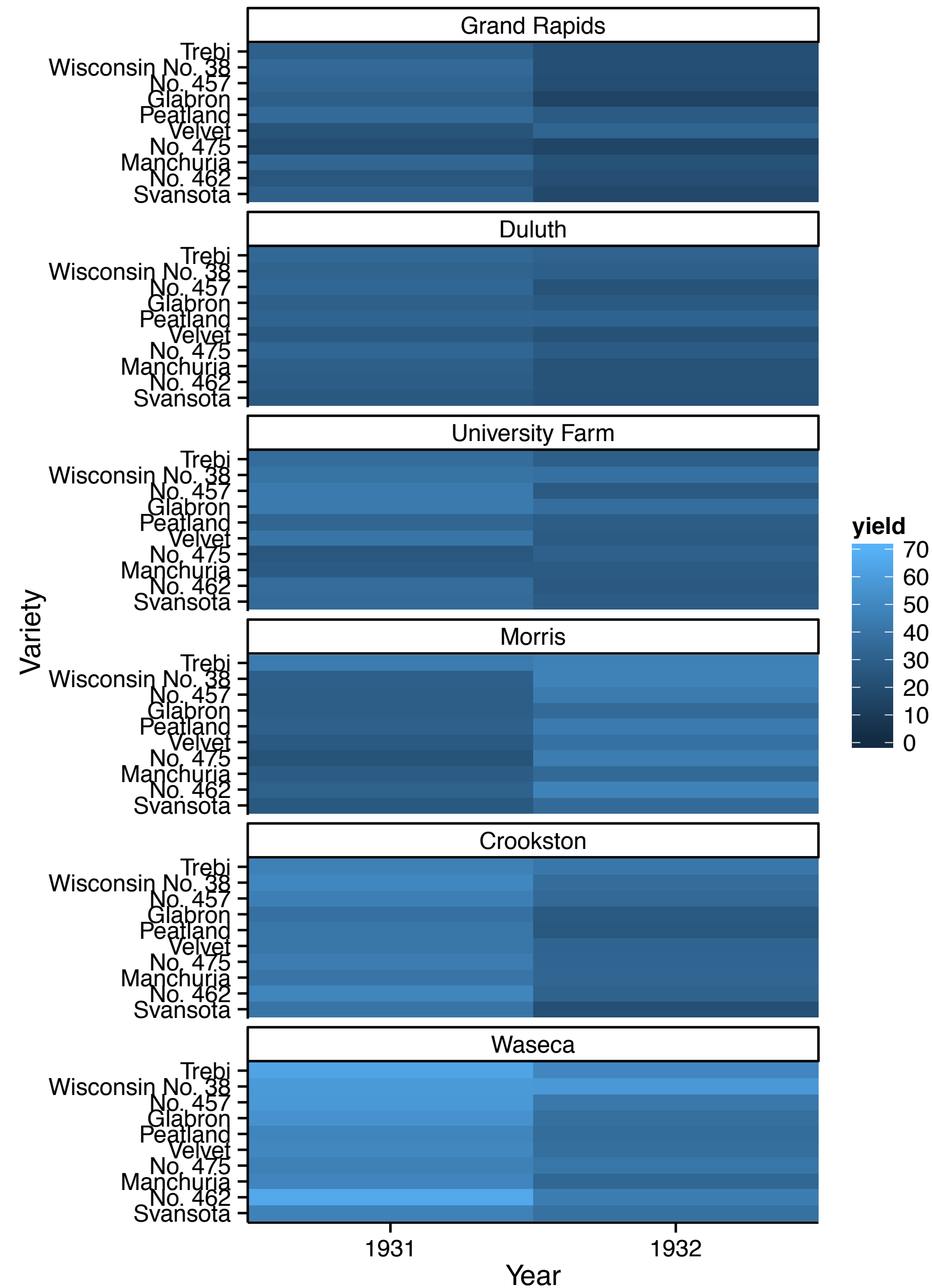
DATA VISUALIZATION WITH GGPLOT2

Heat Maps

barley.s

```
> head(barley.s, 15)
```

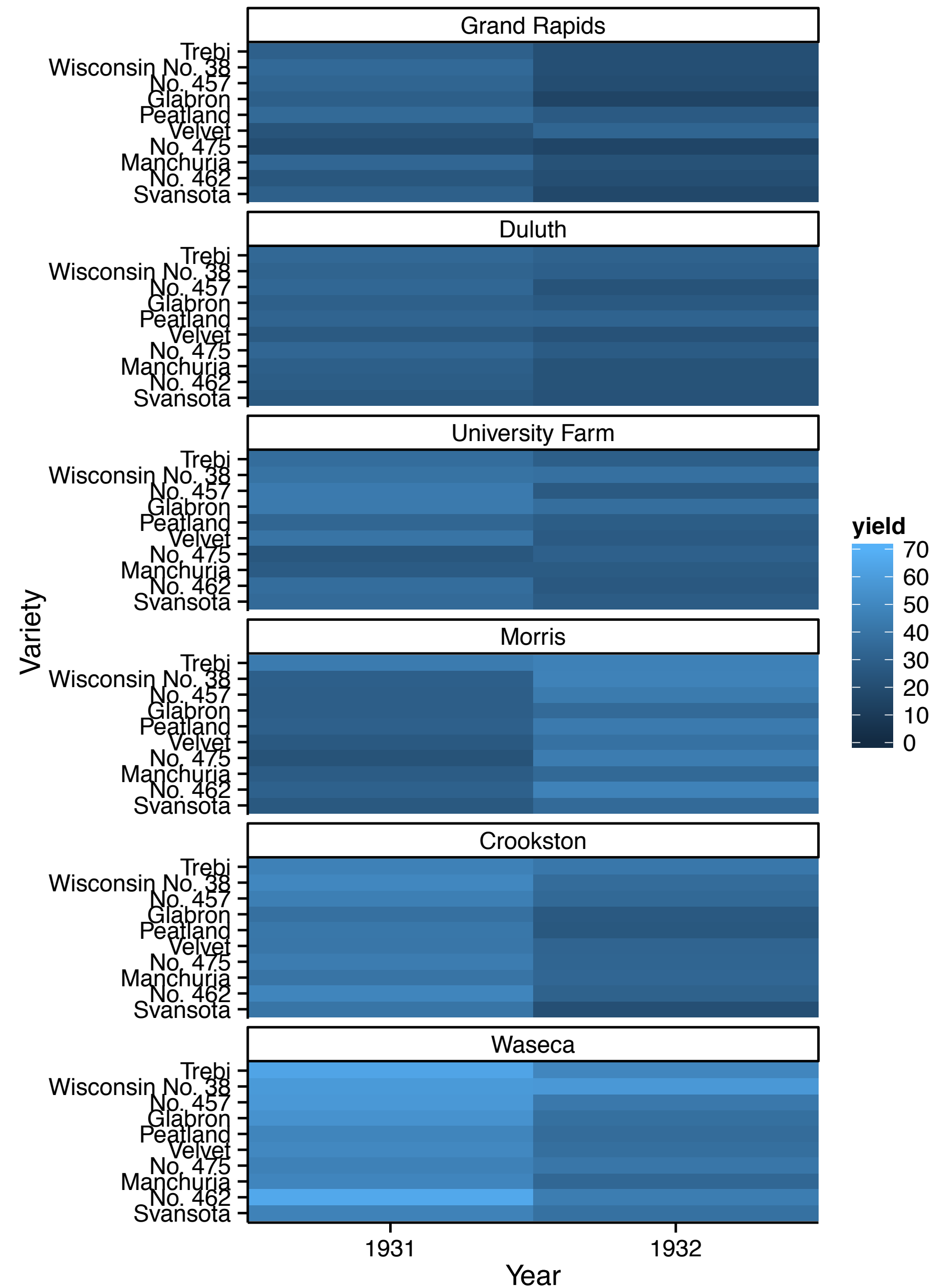
	variety	site	1932	1931
1	Svansota	Grand Rapids	16.63333	29.66667
2	Svansota	Duluth	22.23333	25.70000
3	Svansota	University Farm	27.43334	35.13333
4	Svansota	Morris	35.03333	25.76667
5	Svansota	Crookston	20.63333	40.46667
6	Svansota	Waseca	38.50000	47.33333
7	No. 462	Grand Rapids	19.90000	24.93334
8	No. 462	Duluth	22.50000	28.10000
9	No. 462	University Farm	25.56667	36.60000
10	No. 462	Morris	47.00000	30.36667
11	No. 462	Crookston	30.53333	48.56666
12	No. 462	Waseca	44.70000	65.76670
13	Manchuria	Grand Rapids	22.13333	32.96667
14	Manchuria	Duluth	22.56667	28.96667
15	Manchuria	University Farm	26.90000	27.00000

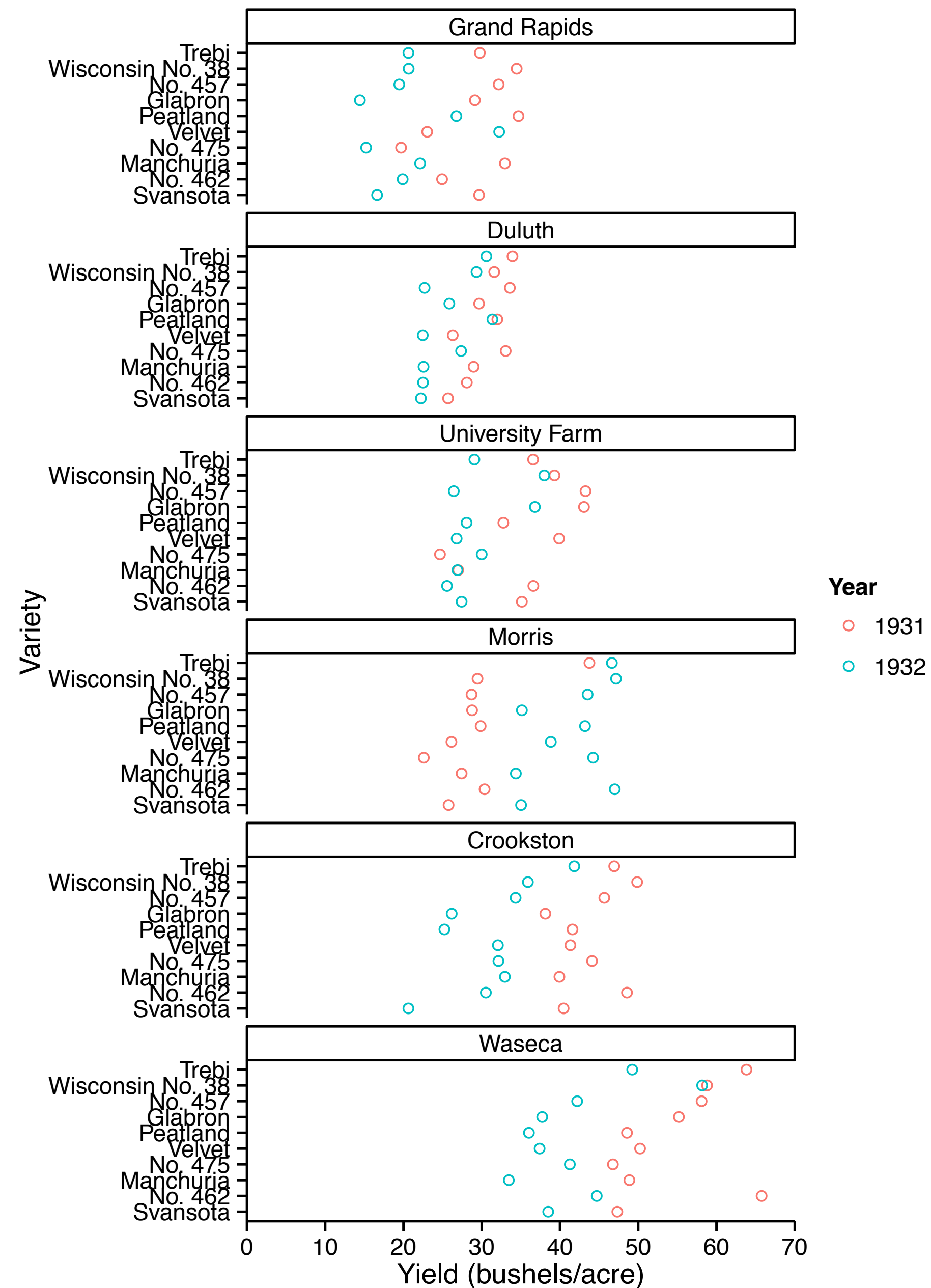


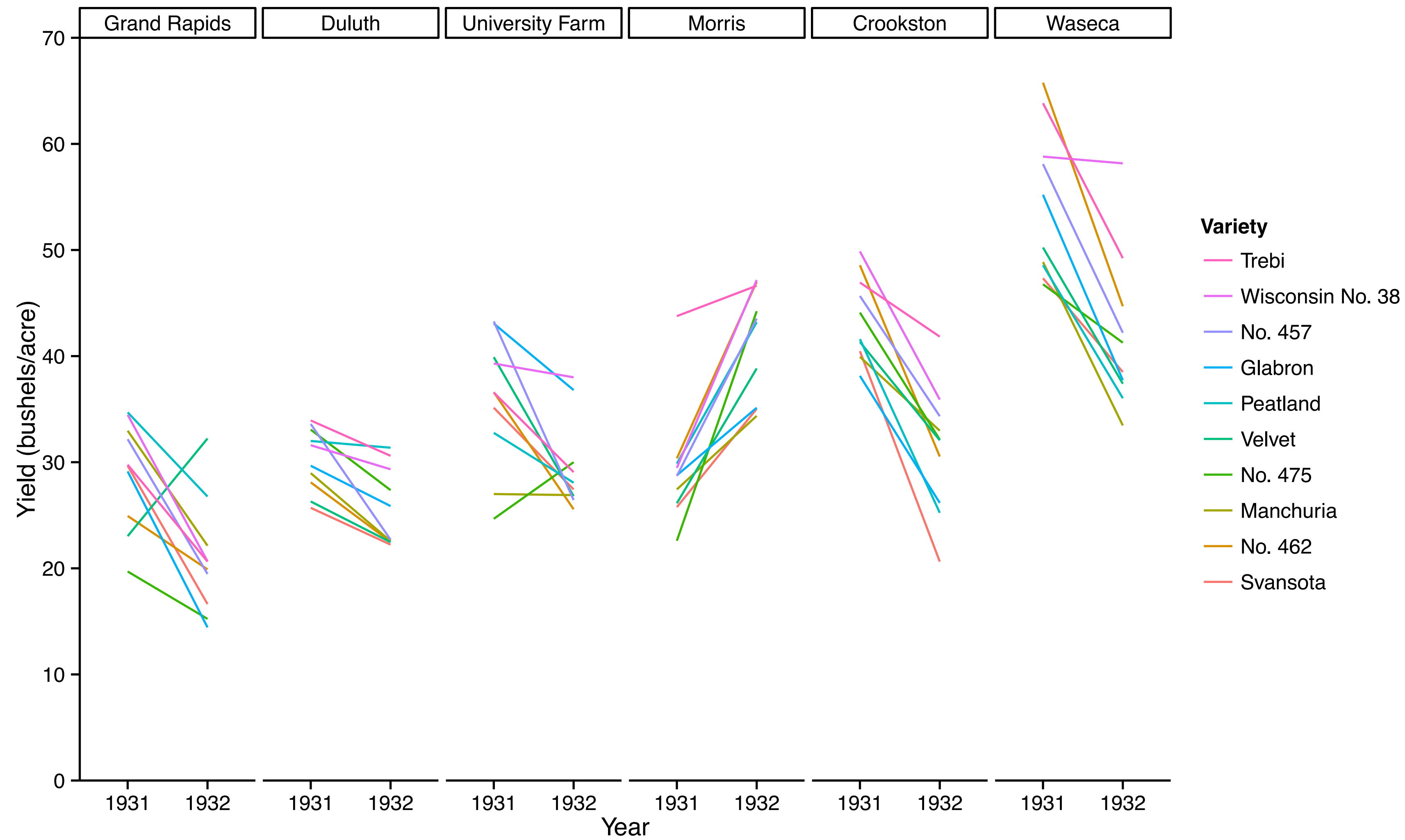
barley

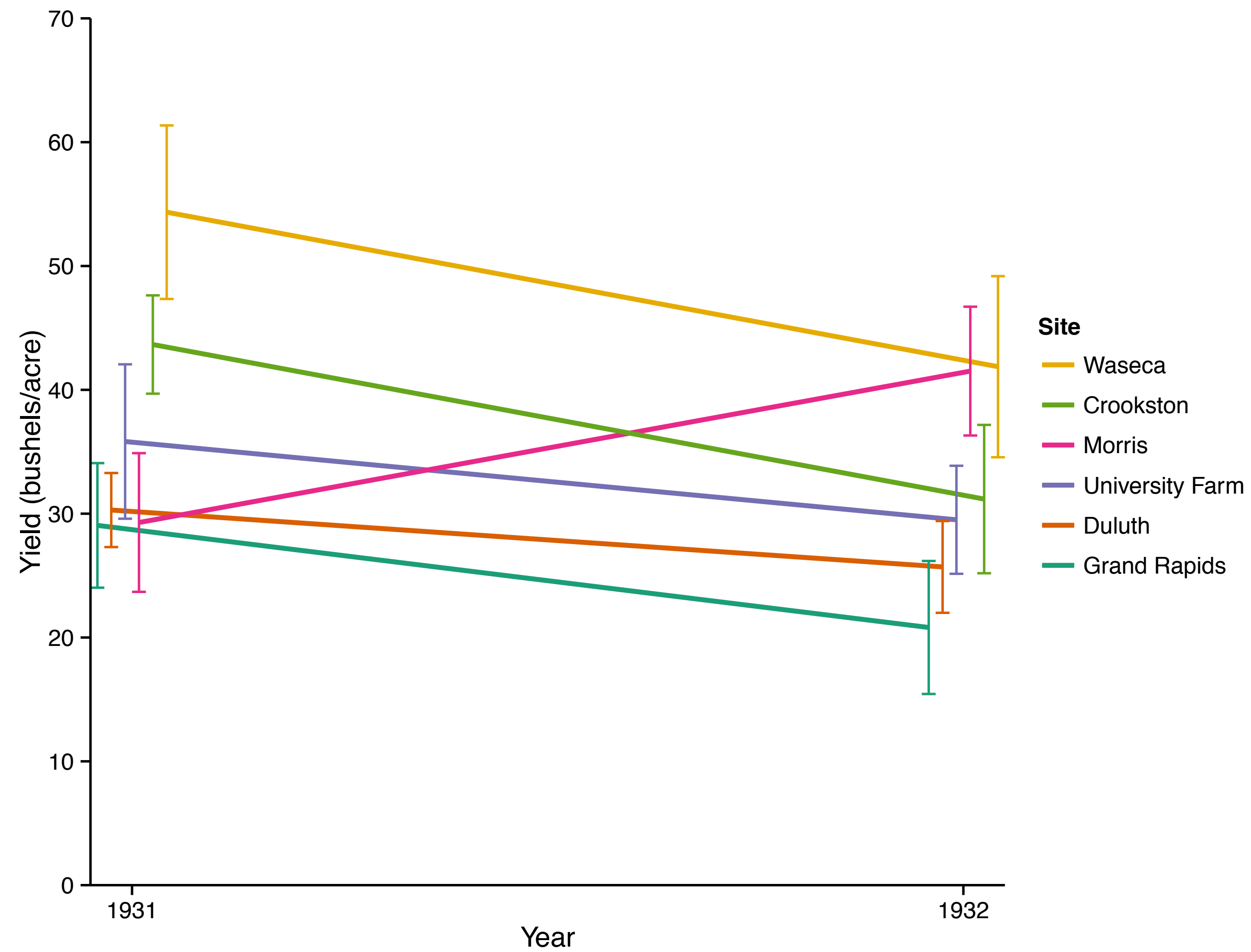
```
> head(barley, 15)
```

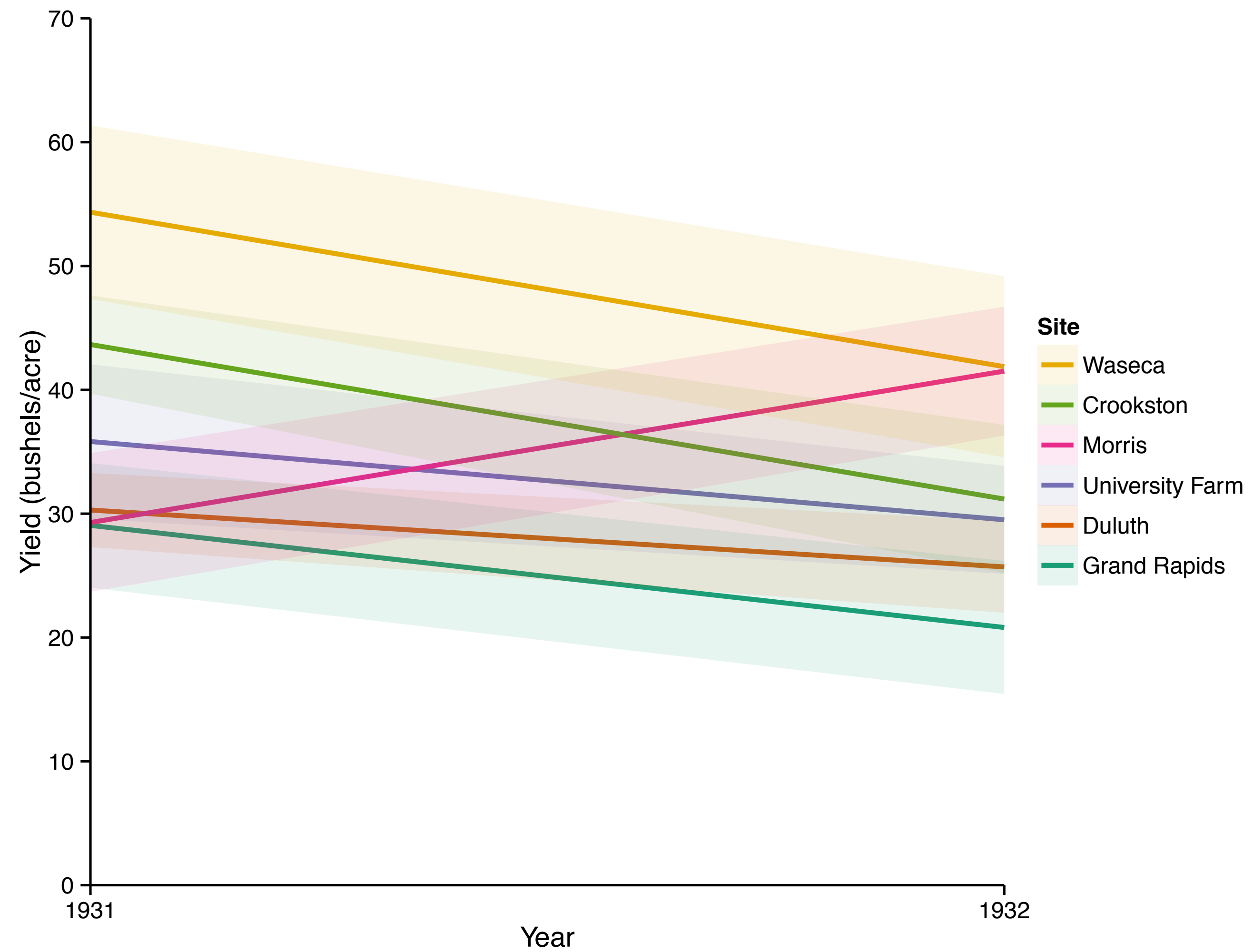
	yield	variety	year	site
1	27.00000	Manchuria	1931	University Farm
2	48.86667	Manchuria	1931	Waseca
3	27.43334	Manchuria	1931	Morris
4	39.93333	Manchuria	1931	Crookston
5	32.96667	Manchuria	1931	Grand Rapids
6	28.96667	Manchuria	1931	Duluth
7	43.06666	Glabron	1931	University Farm
8	55.20000	Glabron	1931	Waseca
9	28.76667	Glabron	1931	Morris
10	38.13333	Glabron	1931	Crookston
11	29.13333	Glabron	1931	Grand Rapids
12	29.66667	Glabron	1931	Duluth
13	35.13333	Svansota	1931	University Farm
14	47.33333	Svansota	1931	Waseca
15	25.76667	Svansota	1931	Morris













DATA VISUALIZATION WITH GGPLOT2

Let's practice!