

Lecture 20 – Grammar of Graphics

Learning Objectives:

6. Learn how to document your work and prepare scientific publications.

6.3 Learn how to plot with ggplot2.

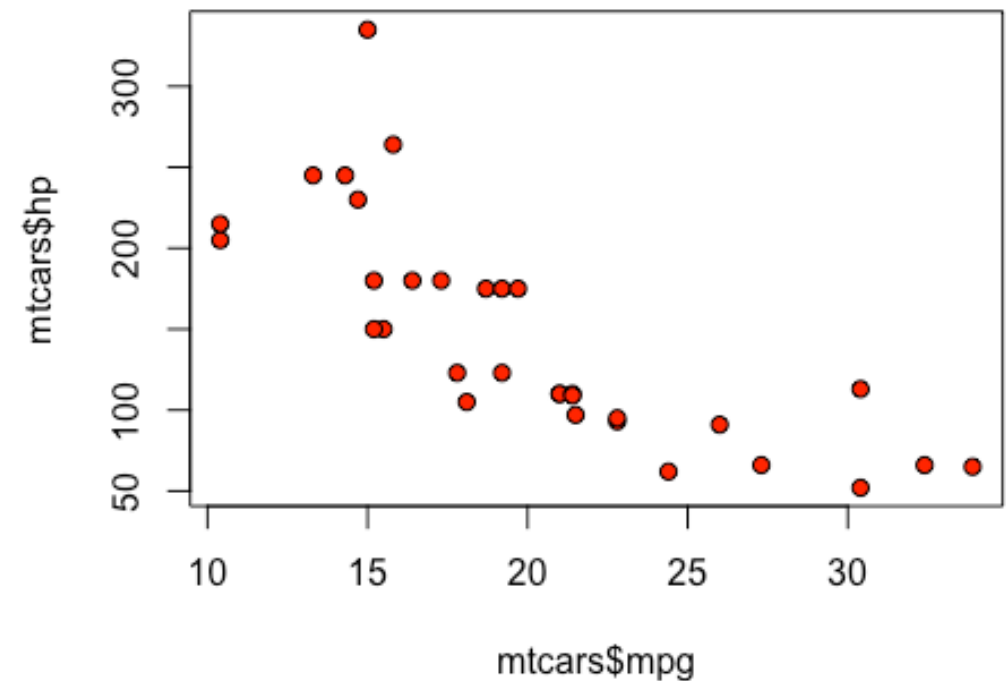
The Plotting Systems of R

- Base graphics and grid

```
plot(x = mtcars$mpg, mtcars$hp,  
     col = "black", bg = "red",  
     pch = 21)
```

Benefits:

- simple and quick
- handles a variety of data types
- not many background calculations

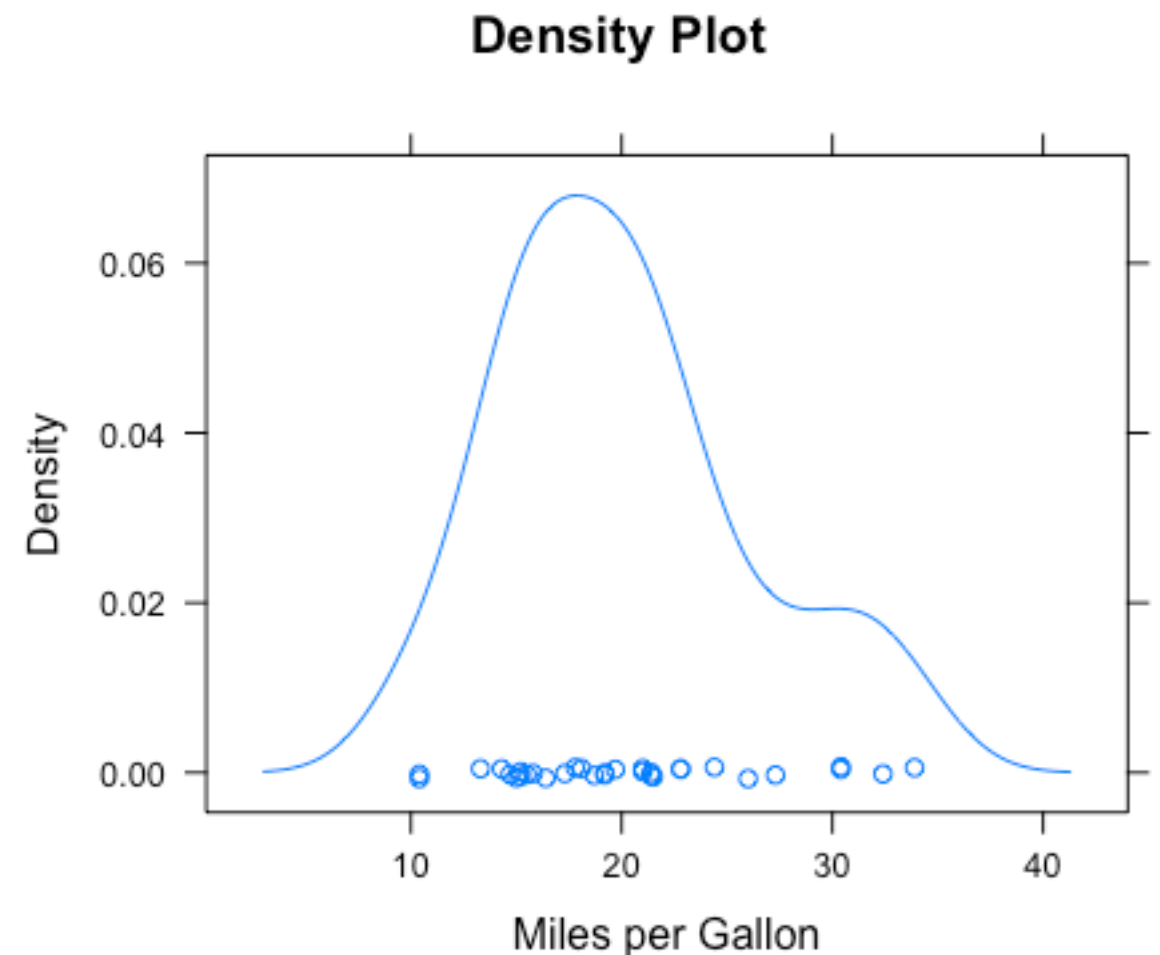


- The Trellis system in lattice

```
densityplot(~mpg,  
            main="Density Plot",  
            xlab="Miles per Gallon")
```

Benefits:

- quick
- many specialized stats plots
- not too many background calculations



The Plotting Systems of R

- Grammar of Graphics `ggplot2`

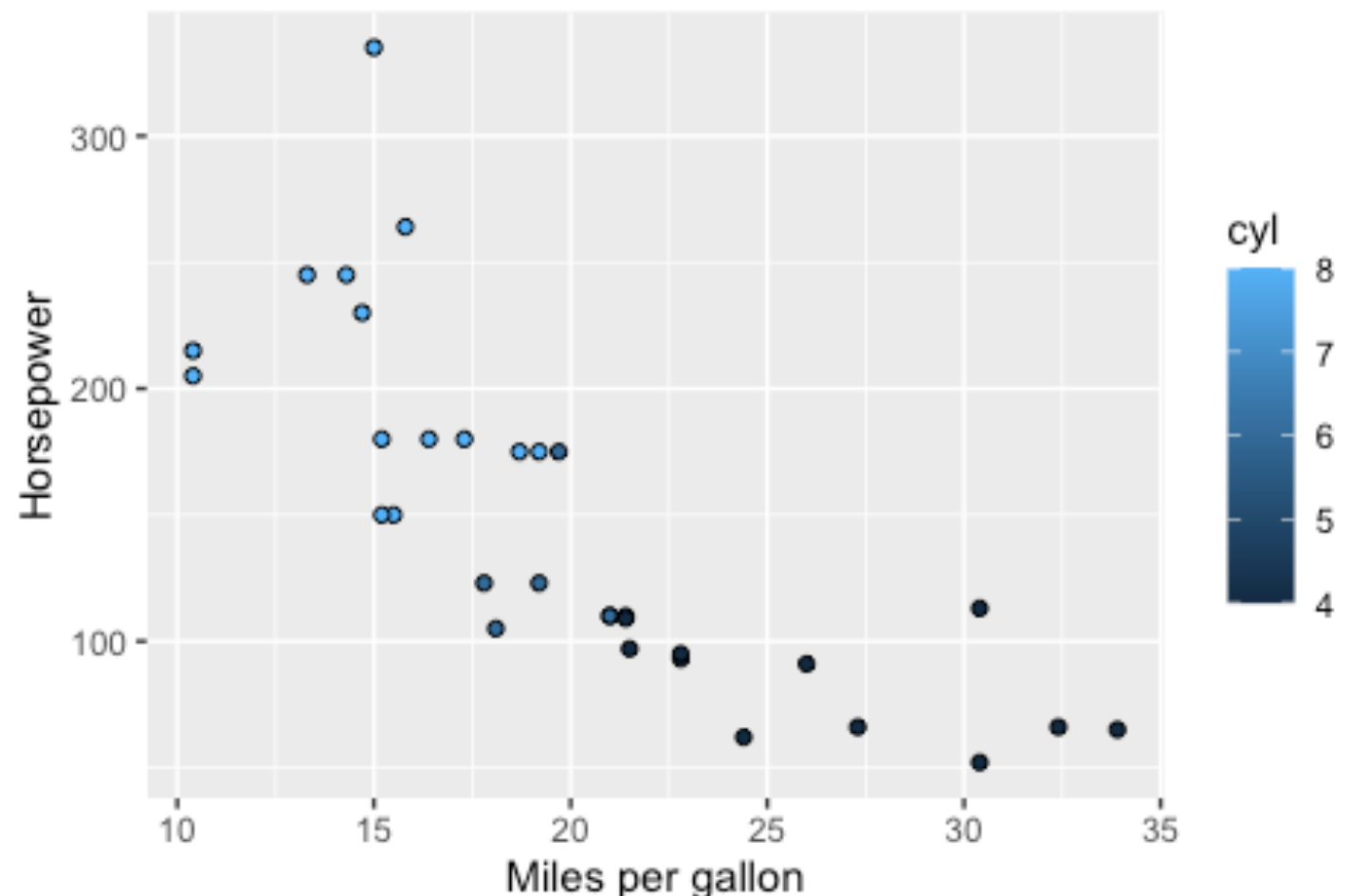
```
ggplot(mtcars, aes(x=mpg,y=hp,fill=cyl)) +  
  geom_point(pch=21) +  
  xlab("Miles per gallon") + ylab("Horsepower")
```

Benefits:

- beautiful visualizations
- many specialized plots
- consistent syntax
- have full control over aesthetics

Drawbacks:

- lots of background calculations
- slow
- only uses data frames in long format



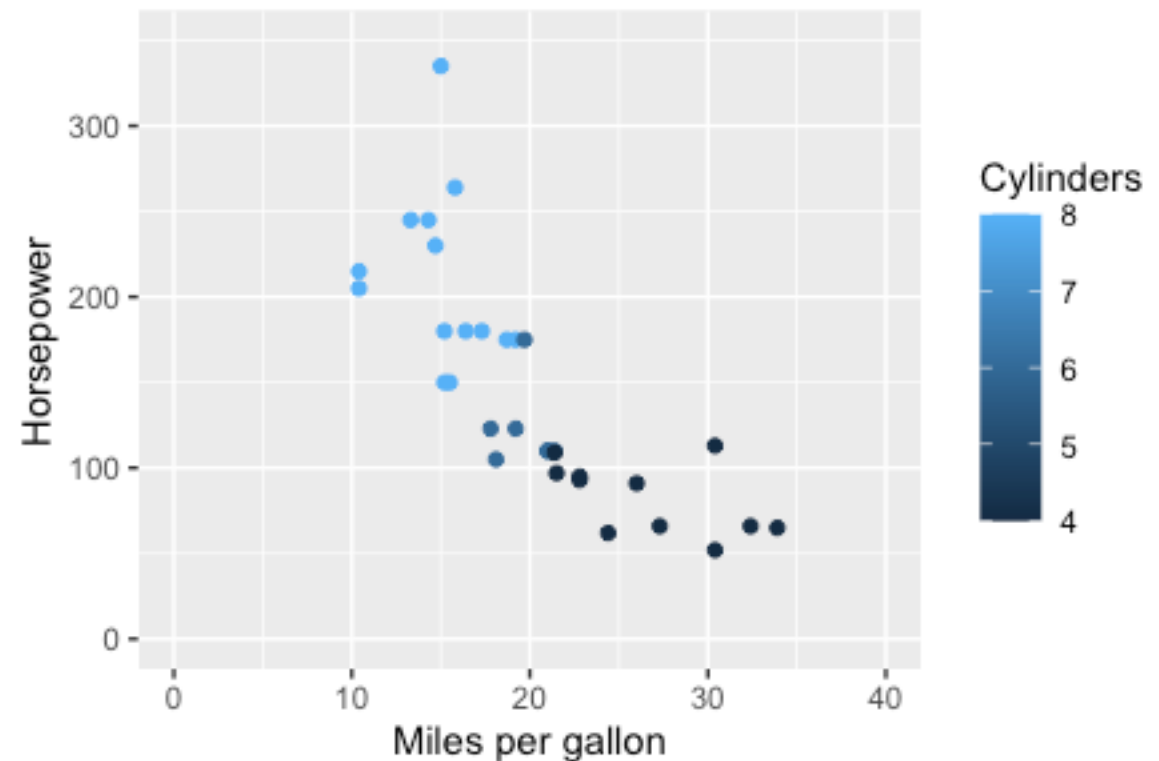
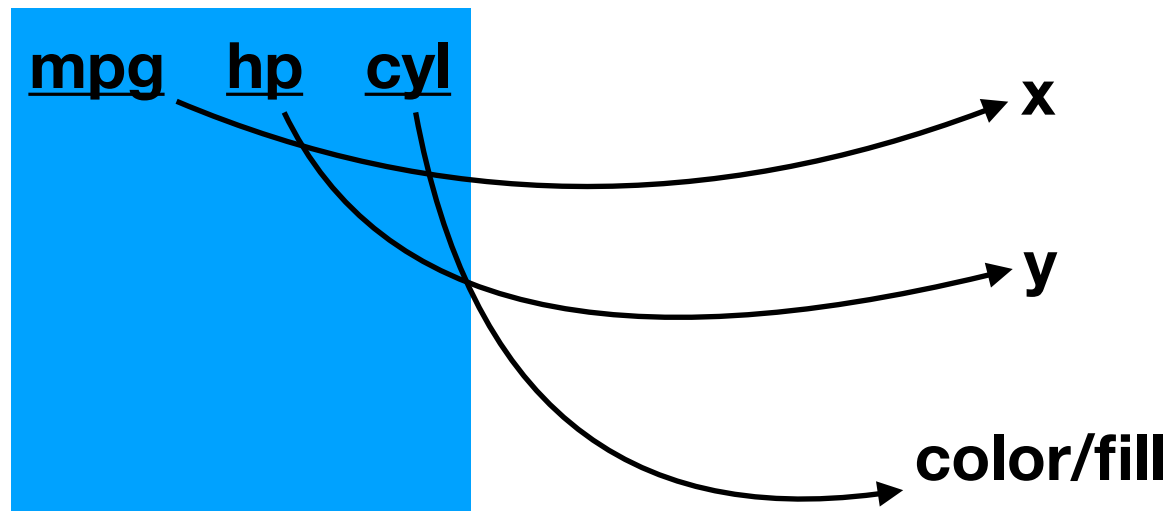
Grammar of Graphics: terminology

Data

Mapping

Aesthetic
attributes

Geometric object



```
ggplot(
```

```
+ geom_point()
```

```
mtcars,
```

```
  aes(x=mpg,  
      y=hp,  
      fill=cyl)
```

Scales

```
+ xlim(0, 40)
```

Guides

```
+ labs(color="Cylinders")
```

Data: wide versus long format

Dataset: Loblolly

Package: reshape2

Wide: Columns represent different measurements

Seed	Year 3	Year 5	Year 10	Year 15	Year 20	Year 25
301	4.51	10.89	28.72	41.75	52.70	60.92
303	4.55	10.92	29.07	42.83	53.88	63.39
305	4.79	11.37	30.21	44.40	52.82	64.10
307	4.81	11.20	28.66	41.66	53.31	63.05

Long: Each row is a unique observation

Seed	Age	Height
301	3	4.51
301	5	10.89
301	10	28.72
301	15	41.74
301	20	52.70
301	25	60.92
303	3	4.55
303	5	10.92
303	10	29.07
303	15	42.38
303	20	53.88
303	25	63.39

Data: wide versus long format

Dataset: Loblolly

Package: reshape2

Wide: Columns represent different measurements

Seed	Year 3	Year 5	Year 10	Year 15	Year 20	Year 25
301	4.51	10.89	28.72	41.75	52.70	60.92
303	4.55	10.92	29.07	42.83	53.88	63.39
305	4.79	11.37	30.21	44.40	52.82	64.10
307	4.81	11.20	28.66	41.66	53.31	63.05

`Loblolly.wide <- dcast(Loblolly, Seed~age,
value.var="height")`

Long: Each row is a unique observation

Seed	Age	Height
301	3	4.51
301	5	10.89
301	10	28.72
301	15	41.74
301	20	52.70
301	25	60.92
303	3	4.55
303	5	10.92
303	10	29.07
303	15	42.38
303	20	53.88
303	25	63.39

Data: wide versus long format

Dataset: Loblolly

Package: reshape2

Wide: Columns represent different measurements

Seed	Year 3	Year 5	Year 10	Year 15	Year 20	Year 25
301	4.51	10.89	28.72	41.75	52.70	60.92
303	4.55	10.92	29.07	42.83	53.88	63.39
305	4.79	11.37	30.21	44.40	52.82	64.10
307	4.81	11.20	28.66	41.66	53.31	63.05

`Loblolly.wide <- dcast(Loblolly, Seed~age,
value.var="height")`

`Loblolly.long <- melt(Loblolly.wide)`

Long: Each row is a unique observation

Seed	Age	Height
301	3	4.51
301	5	10.89
301	10	28.72
301	15	41.74
301	20	52.70
301	25	60.92
303	3	4.55
303	5	10.92
303	10	29.07
303	15	42.38
303	20	53.88
303	25	63.39

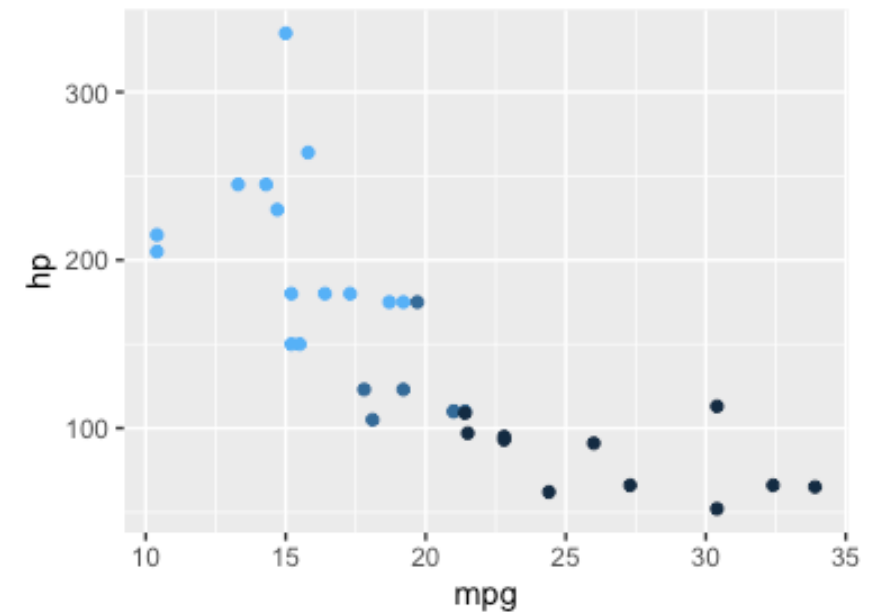
Mapping aesthetics based on data

- Mapping aesthetics that change with data must be done with `aes ()`

```
ggplot(mtcars,  
      aes(x = mpg, y = hp,  
          color = cyl)) +  
geom_point()
```



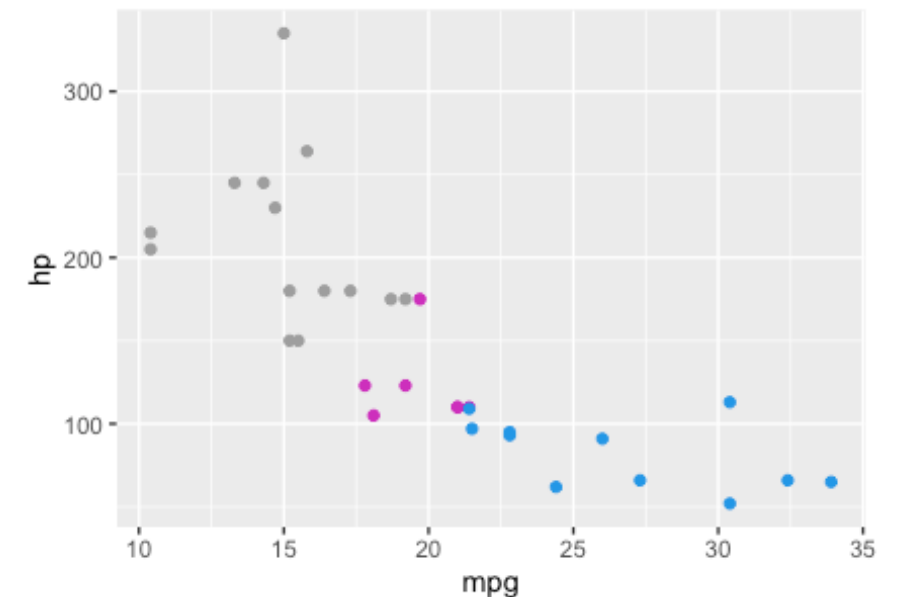
Maps correctly



```
ggplot(mtcars,  
      aes(x = mpg, y = hp)) +  
geom_point(color = cyl)
```



Maps incorrectly



```
ggplot(mtcars,  
      aes(x = mpg, y = hp)) +  
geom_point(aes(color = cyl))
```

Maps correctly

- Changing non-mapped aesthetics can be done anywhere

Geometric objects

Scatter plot – `geom_point()`

Line plot – `geom_line()`

Box plot – `geom_boxplot()`

Violin plot – `geom_violin()`

Bar plot – `geom_bar()`

Contour plot – `geom_contour()`

Density plot – `geom_density()`

Plot a map – `geom_map()`

Rectangles – `geom_raster()`
`geom_tile()`

**Quantile-quantile
plot** – `geom_qq_line()`

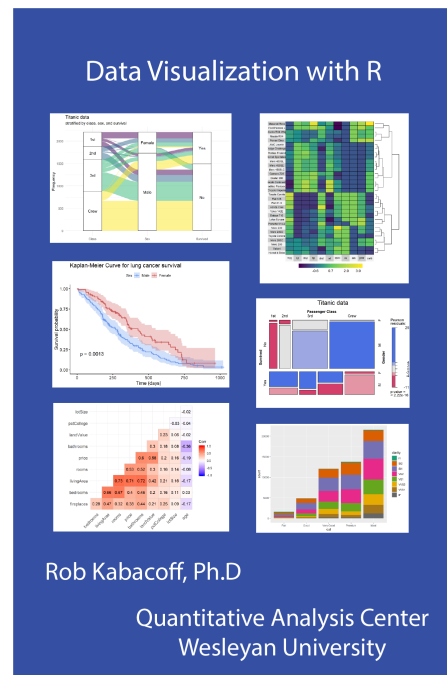
Stacked dot plot – `geom_dotplot()`

Histogram – `geom_histogram()`

**Choose the correct plot
for your data!**

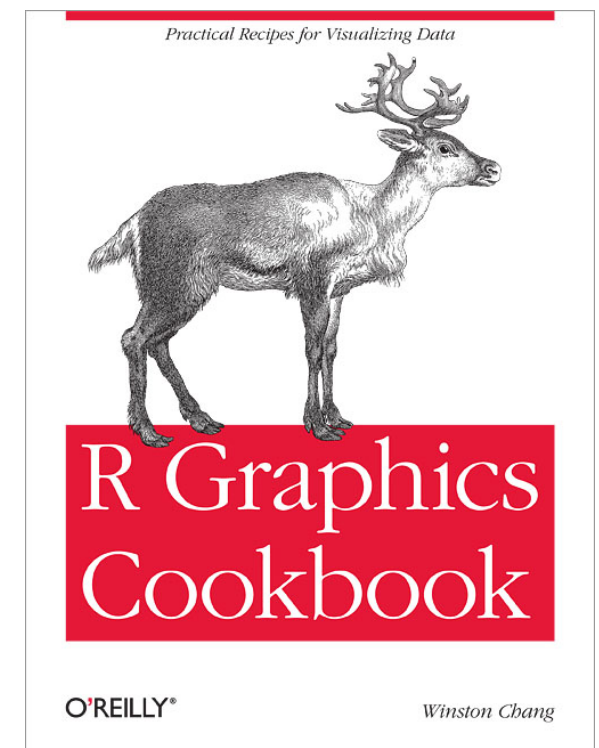
Excellent Resources for ggplot2

Tidyverse Reference Guide – <https://ggplot2.tidyverse.org/reference/>



Data Visualization with R – <https://rkabacoff.github.io/datavis/>

The R Graphics Cookbook – <https://r-graphics.org/>



Additional Resources

<https://www.stat.auckland.ac.nz/~ihaka/787/lectures-trellis.pdf> –

The Trellis system in lattice (PDF lecture slides)

http://www.cookbook-r.com/Manipulating_data/

Converting data between wide and long format/ –

Converting between long and wide format with reshape2, tidyr, and base R