Rotman

INTRO TO R - VISUALIZATION

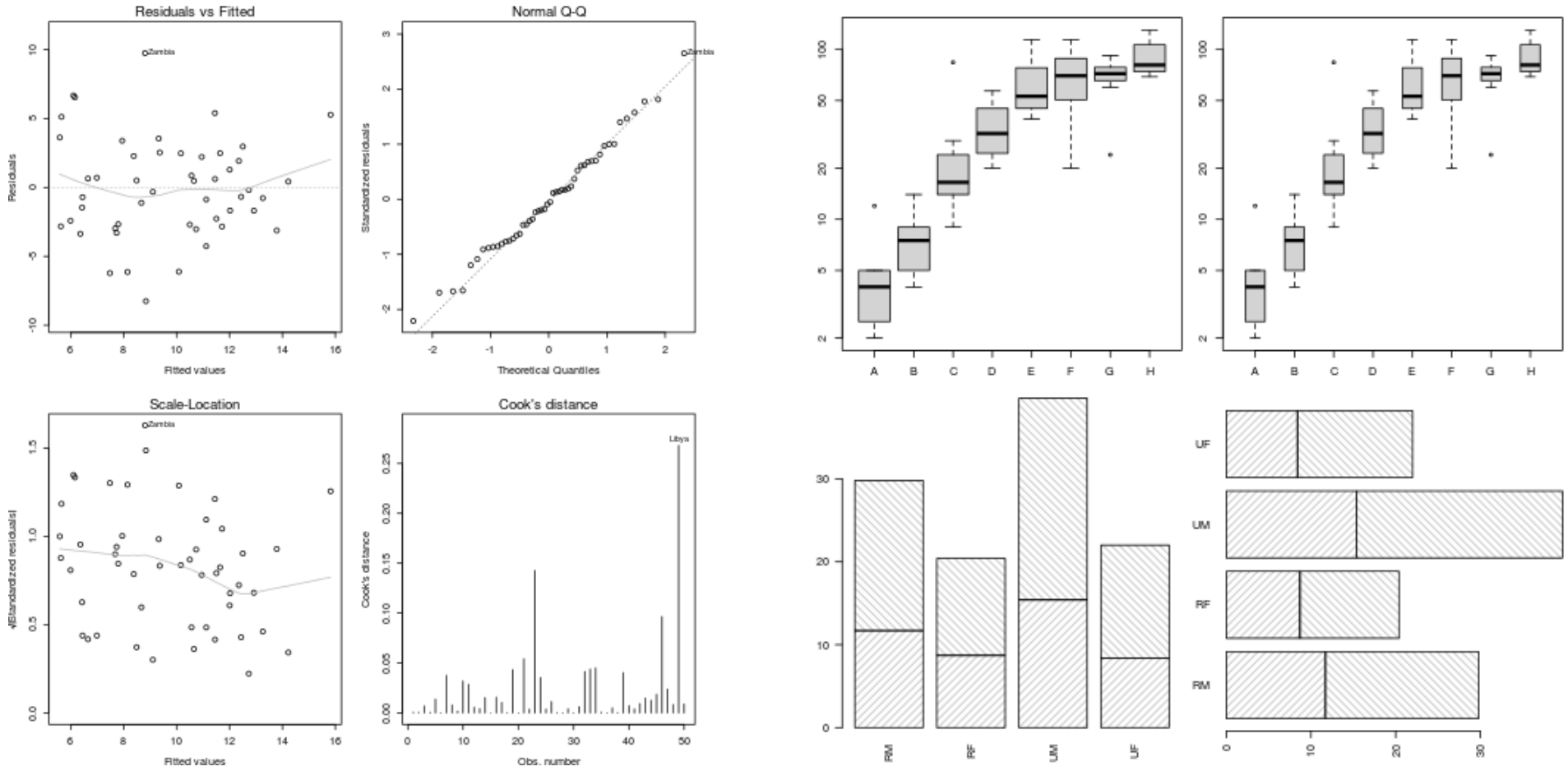
R Workshop - 3

February 18, 2020 Prepared by Jay Cao / TDMDAL

Website: https://tdmdal.github.io/r-tutorial-201920-winter/

Rotman School of Management
UNIVERSITY OF TORONTO

R Graphics – Base plots (examples)

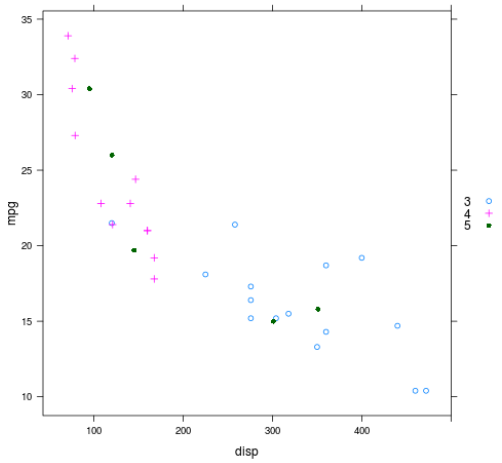


https://www.stat.auckland.ac.nz/~paul/RG3e/chapter2.html

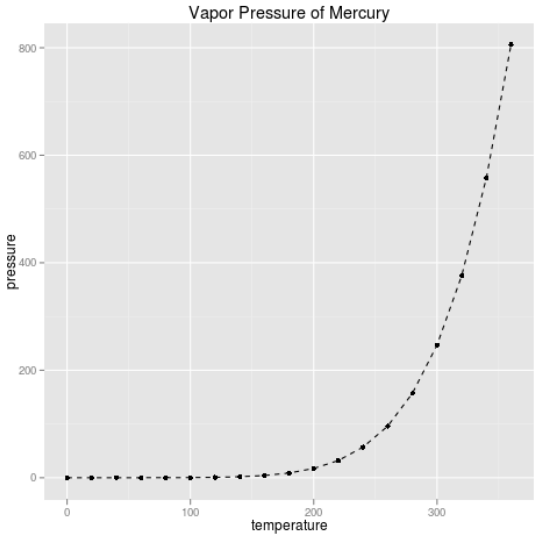
R Graphics – Two Main Plotting Systems

- System?

- R package: lattice
  - implements Trellis system by William Cleveland:



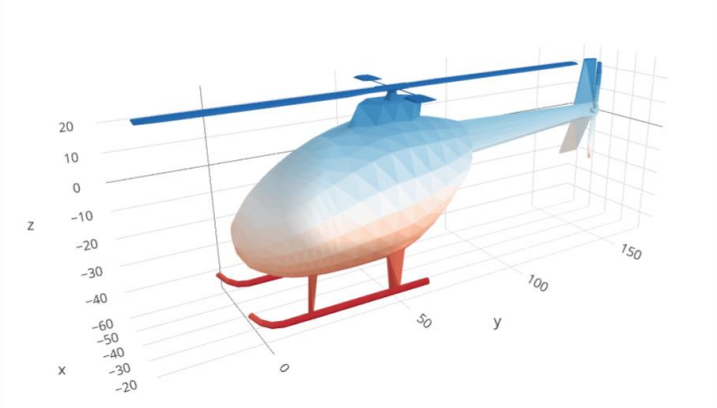
- R package: ggplot2
  - implements "A Grammar of Graphics" by Leland Wilkinson
  - **our focus today**



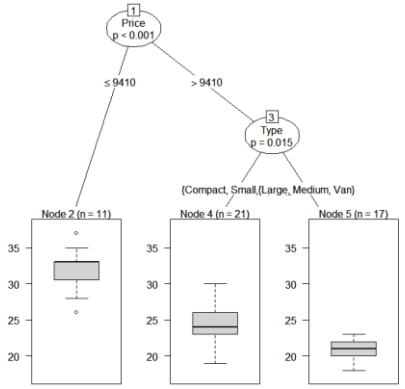
https://www.stat.auckland.ac.nz/~paul/RG3e/chapter4.html
https://www.stat.auckland.ac.nz/~paul/RG3e/chapter5.html

Other Specialized Plots

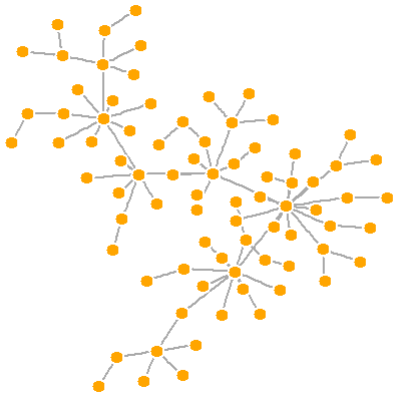
- Graphic functions provided by specialized packages
  - Based on R primitive graphical engines like grid (eg. plot() in party, igraph)
  - Following a plotting system (eg. ggmap, tmap, gganimate, plotly, etc.)
  - Wrapper of plotting tools in another languages (ex. leaflet, grViz() in DiagrammeR)



3D tri-surface interactive plot using the plotly package https://plot.ly/r/trisurf/



Decision tree plot using party package https://www.statmethods.net/advstats/cart.html



Network plot using igraph package http://kateto.net/networks-r-igraph

Other Specialized Plots – One More Example

https://www.quantmod.com/examples/charting/

ggplot2

- Based on the Grammar of Graphics

- Basic idea: you can build any graph from the same components
  - Data
  - Coordinate system
  - Geoms – visual marks that represent data points

- A layer-by-layer approach

http://amzn.to/2ef1eWp
http://vita.had.co.nz/papers/layered-grammar.html

ggplot() – “base layer”

data

```
p <- ggplot(df, aes(x, y, other_aesthetics))
```

https://ggplot2.tidyverse.org/reference/ggplot.html

ggplot() – “base layer”

mapping: linking variables in the data to aesthetic elements in the plot

data

```
p <- ggplot(df, aes(x, y, other_aesthetics))
```

https://ggplot2.tidyverse.org/reference/ggplot.html

ggplot() – “base layer”

mapping: linking variables in the data to aesthetic elements in the plot

data

```
p <- ggplot(df, aes(x, y, other_aesthetics))
```

(x, y) coordinates mapping

color-, size- mapping, etc.

https://ggplot2.tidyverse.org/reference/ggplot.html

ggplot() – Add Other Layers

mapping: linking variables in the data to aesthetic elements in the plot

data

```
p <- ggplot(df, aes(x, y, other_aesthetics)) +
  another_layer +
  another_layer +
  …
```

what to plot (geom-, scale-functions, etc.): point, line, label, etc.

- If data and mapping are not specified in the base layer, they must be supplied in each layer added to the plot

https://ggplot2.tidyverse.org/reference/ggplot.html

ggplot() – “base layer” / example

```
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       2007    43.8 31889923      975.
## 2 Albania     Europe     2007    76.4  3600523     5937.
## 3 Algeria     Africa     2007    72.3 33333216     6223.
```

```
p <- ggplot(df, aes(x = gdpPercap, y = lifeExp))
```

ggplot() – geom layers (eg. geom_point )

layer specific data and mapping
If not specified, inherit from base layer

```
p + geom_point(DATA, MAPPING, STAT, POSITION, …)
```

https://ggplot2.tidyverse.org/reference/geom_point.html

ggplot() – geom layers (eg. geom_point )

layer specific data and mapping
If not specified, inherit from base layer

```
p + geom_point(DATA, MAPPING, STAT, POSITION, …)
```

statistical transformation &
position adjustment
e.g. `position = “jitter"`

https://ggplot2.tidyverse.org/reference/geom_point.html

ggplot() – geom layers (eg. geom_point )

other arguments:
e.g. `color = "red"`,
`alpha = 0.5`, etc.

layer specific data and mapping
If not specified, inherit from base layer

```
p + geom_point(DATA, MAPPING, STAT, POSITION, …)
```

statistical transformation &
position adjustment
e.g. `position = “jitter"`

https://ggplot2.tidyverse.org/reference/geom_point.html

ggplot() – geom_point layer / example

```
p +
  geom_point(aes(size = pop,
    color = continent),
    alpha = 0.5)
```

ggplot() – example (diamond data)

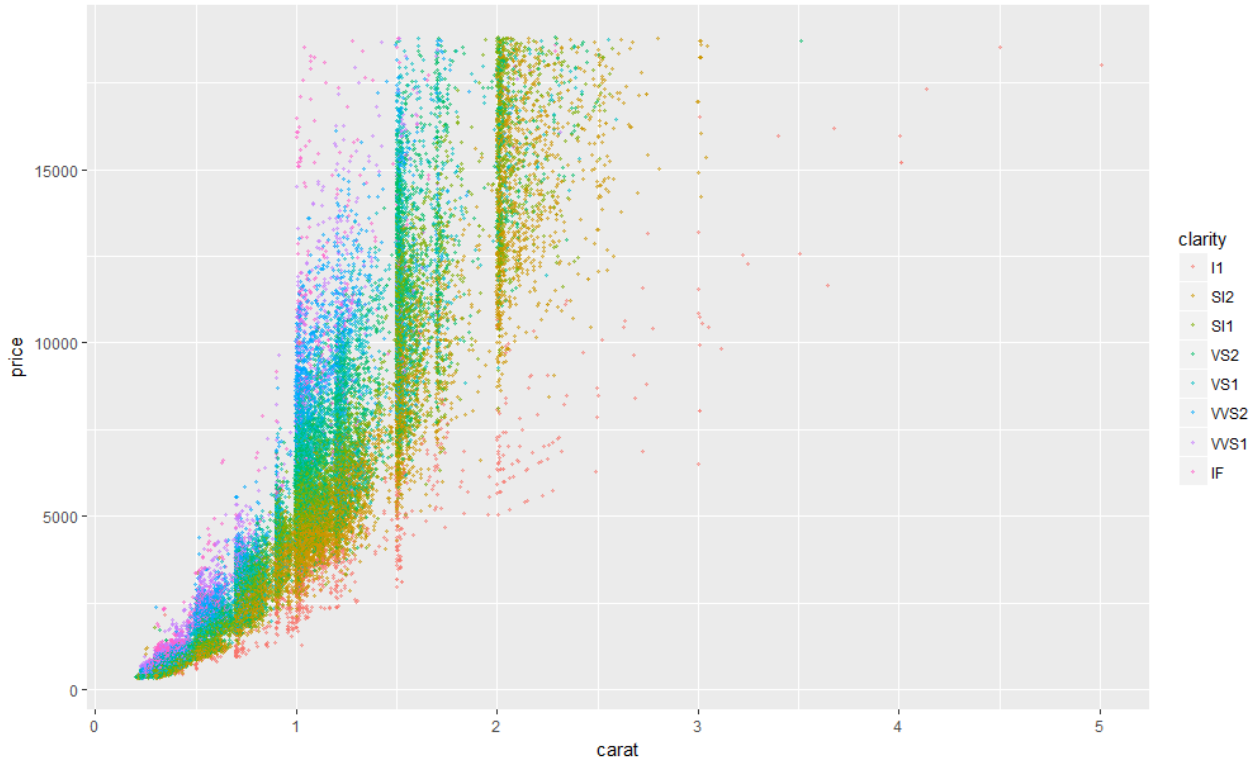
```
## # A tibble: 6 x 10
##   carat cut       color clarity depth table price     x     y     z
##   <dbl> <ord>     <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1 0.23  Ideal     E     SI2      61.5    55   326  3.95  3.98  2.43
## 2 0.21  Premium   E     SI1      59.8    61   326  3.89  3.84  2.31
## 3 0.23  Good      E     VS1      56.9    65   327  4.05  4.07  2.31
## 4 0.290 Premium   I     VS2      62.4    58   334  4.2   4.23  2.63
## 5 0.31  Good      J     SI2      63.3    58   335  4.34  4.35  2.75
## 6 0.24  Very Good J     VVS2     62.8    57   336  3.94  3.96  2.48
```

ggplot() – example (layer 1)

```
ggplot(data = diamonds, aes(carat, price)) +
```

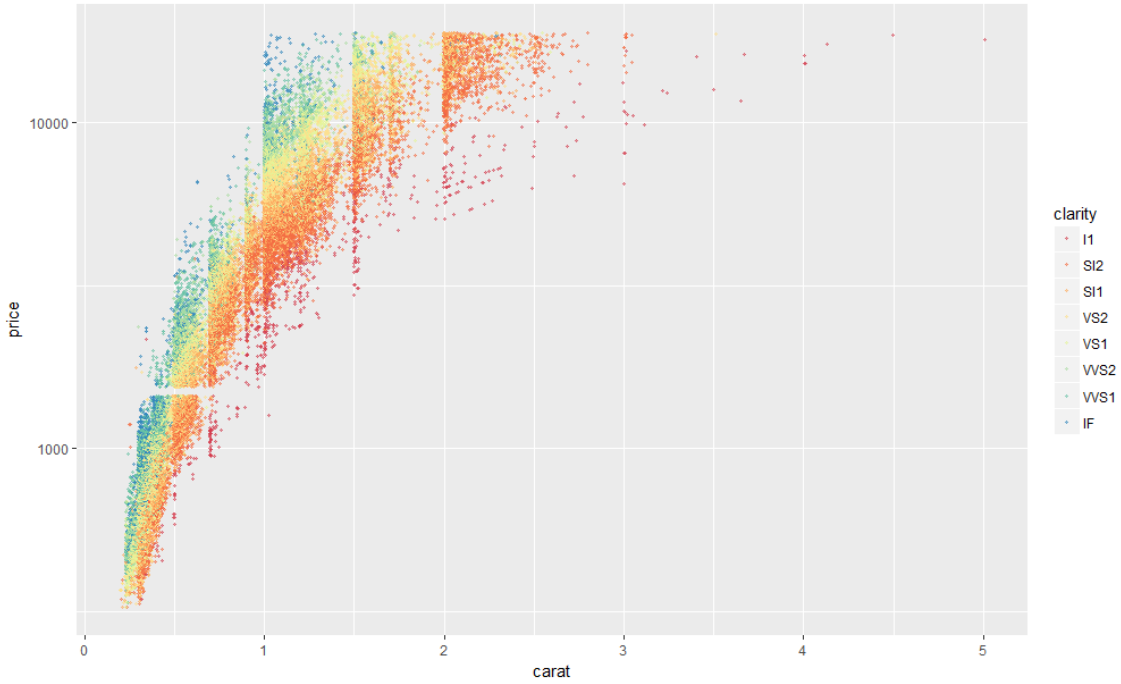
ggplot() – example (layer 2)

```
ggplot(data = diamonds, aes(carat, price)) +
  geom_point(aes(colour = clarity),
    position = "jitter",
    alpha = 0.5,
    size = 0.8) +
```



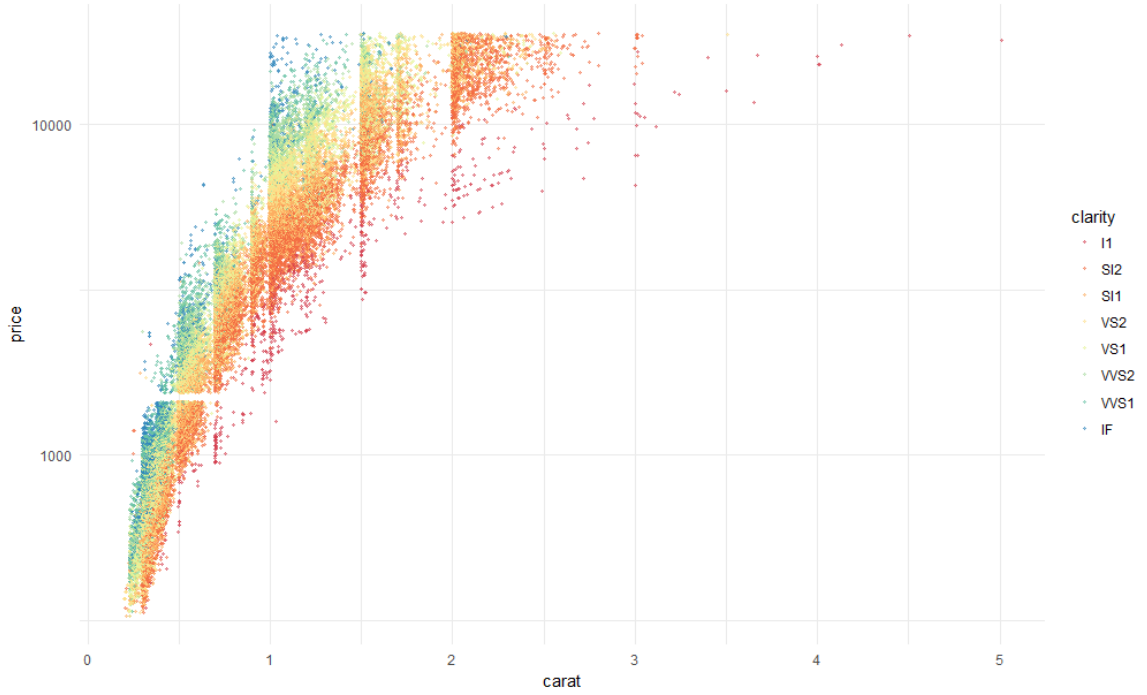
ggplot() – example (layer 3 & layer 4)

```
ggplot(data = diamonds, aes(carat, price)) +
  geom_point(aes(colour = clarity),
    position = "jitter",
    alpha = 0.5,
    size = 0.8) +
  scale_y_continuous(trans = "log10") +
  scale_color_brewer(palette = "Spectral")
```



ggplot() – example (layer 5)

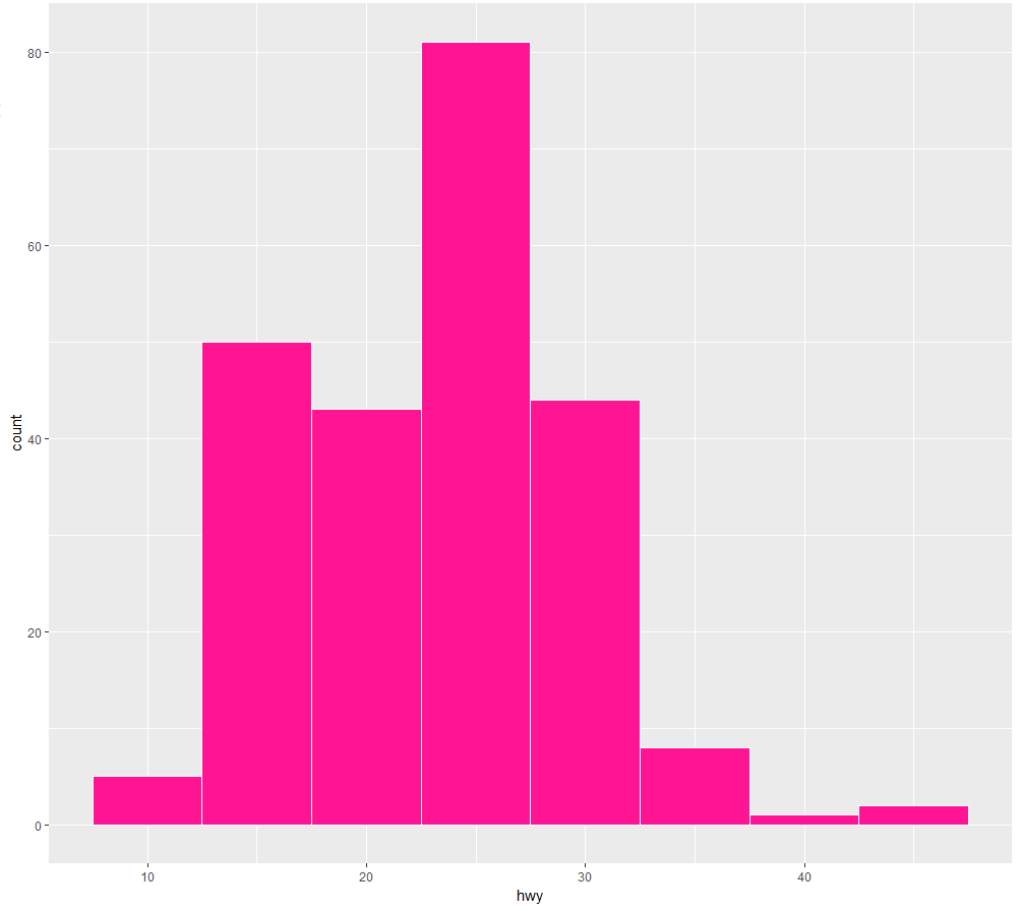
```
ggplot(data = diamonds, aes(carat, price)) +
  geom_point(aes(colour = clarity),
    position = "jitter",
    alpha=0.5,
    size = 0.8) +
  scale_y_continuous(trans = "log10") +
  scale_color_brewer(palette = "Spectral") +
  theme_minimal()
```



ggplot() – geom_histogram / example

```
# A tibble: 234 x 11
   manufacturer model      displ  year   cyl trans      drv     cty   hwy fl    c
   <chr>        <chr>      <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <
 1 audi         a4           1.80  1999     4 auto(l5)   f        18    29 p     c
 2 audi         a4           1.80  1999     4 manual(m5) f        21    29 p     c
 3 audi         a4           2.00  2008     4 manual(m6) f        20    31 p     c
```

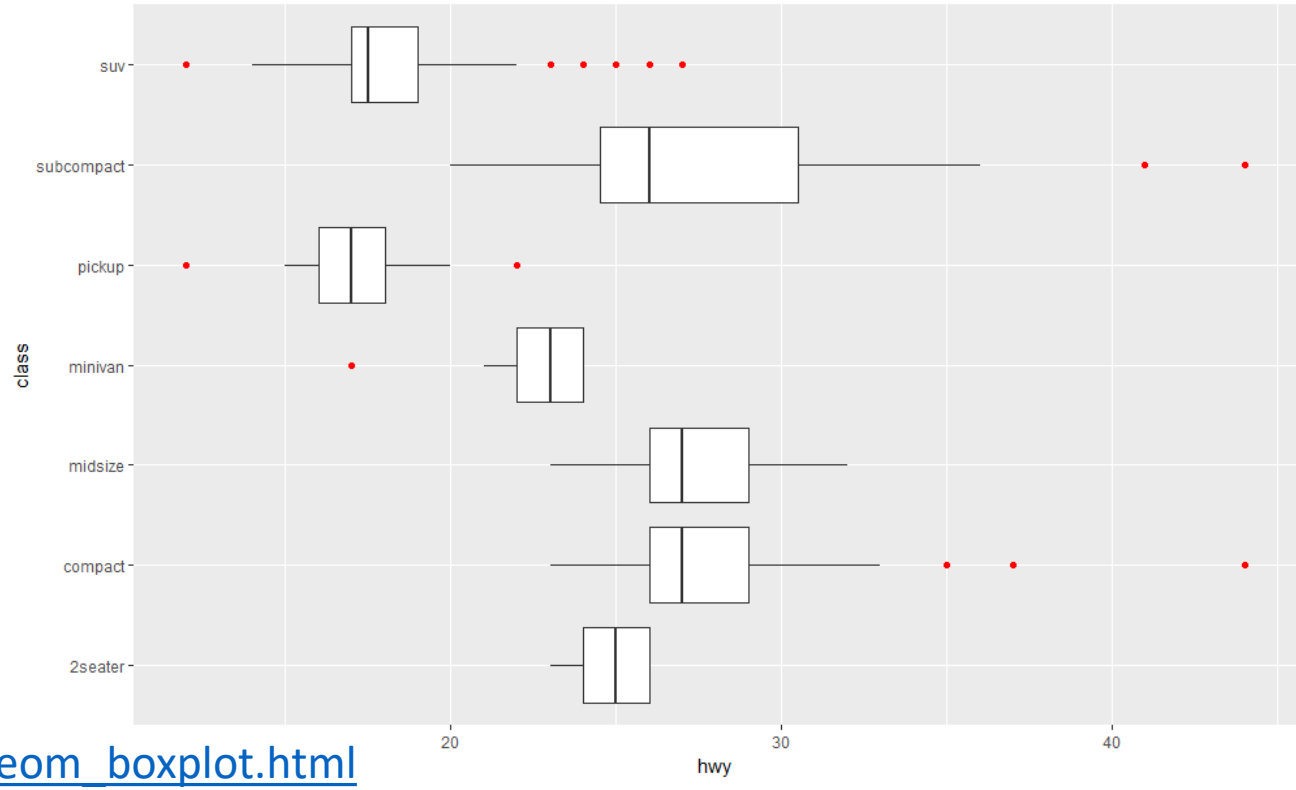
```
ggplot(mpg, aes(x = hwy)) +
  geom_histogram(binwidth=5,
    color = "white",
    fill = "deeppink")
```



https://ggplot2.tidyverse.org/reference/geom_histogram.html

ggplot() – geom_boxplot / example

```
ggplot(mpg, aes(class, hwy)) +
  geom_boxplot(outlier.colour = "red") +
  coord_flip()
```



https://ggplot2.tidyverse.org/reference/geom_boxplot.html

Learning Resources

- ggplot2: Elegant Graphics for Data Analysis (3rd ed.; work in progress)
- R Graphics Cookbook (2nd ed.)