**ggplot2**

**Grammar of Graphics**

2 principles

1. Graphics = distinct layers of grammatical elements
2. Meaningful plots through aesthetic mapping

|  |  |
| --- | --- |
|  |  |

ggplot(mtcars, aes(x = wt, y = mpg)) + geom\_point()

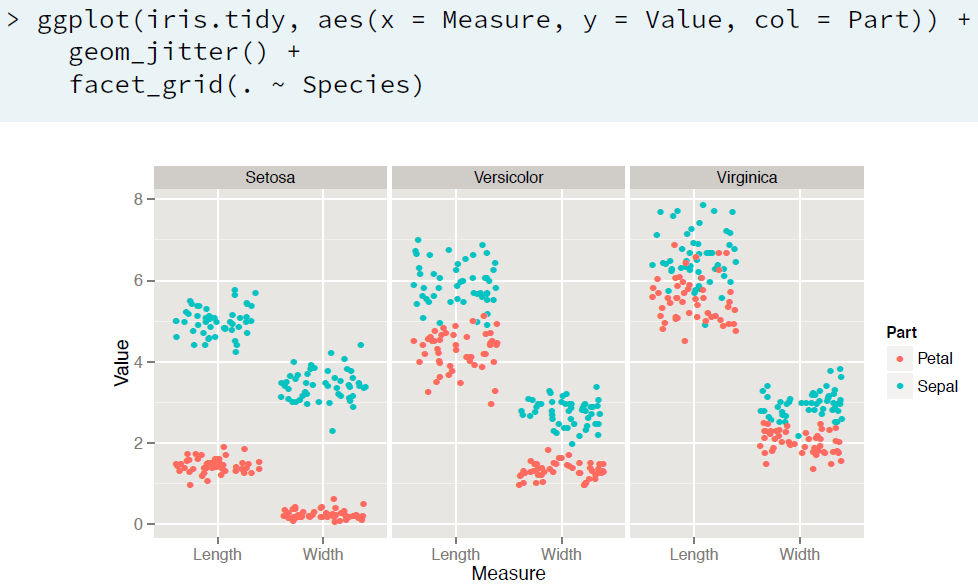
ggplot(mtcars, aes(x = wt, y = mpg, **color** = disp)) + geom\_point()

ggplot(mtcars, aes(x = wt, y = mpg, **size** = disp)) + geom\_point()

|  |  |
| --- | --- |
|  |  |

**Data**

iris.wide (pg15) & iris.tidy (pg23) & facet\_grid()



**Aesthetics**

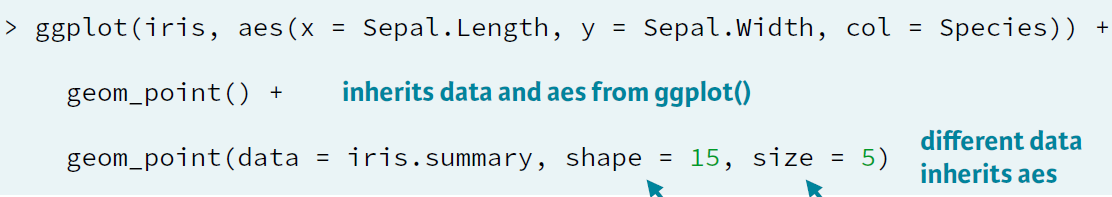
* Column can be mapped onto visible ***aesthetic***
* Aesthetics in aes(), ***attributes*** in geom\_(col=”red”)
* aes() can also be called in geom\_(), but done usually when you want to include multiple data sources
* ggplot(mtcars, aes(x=wt, y=mpg, fill=cyl, col=am)) + geom\_point(**shape**=21, **size**=4, **alpha**=0.6)
  + aes has to be associated with columns
  + attributes are given along with geom\_\*() and don’t have columns associated with them
* ggplot(mtcars, aes(x = wt, y = mpg, fill = cyl, **label**=rownames(mtcars))) + geom\_**text**(color='red')
* **Modifying Aesthetics**
  + geom\_bar(**postion**=”< stack, fill, dodge, … >”)
  + **scale\_\*** functions
    - scale\_x/y\_continuous/discrete(“title”, limits, breaks, expand, ….)
  + **labs**(x,y,col, …)

|  |  |
| --- | --- |
| Aesthetics for Continuous Variables | Aesthetics for Categorical Variables |

**Geometry Layer**

**Scatter Plots**: geom\_point()

* aes() inside geom\_\*() is same as aes() in ggplot()



* ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, col = Species)) + geom\_**jitter**(**shape**=1, **alpha**=0.6)
  + to visualize the density, use jitter along with alpha & shape(hollow shapes preferred)

**Bar Plots:**

* **Histogram**: geom\_histogram() : x-axis : continuous variables
  + ggplot(df, aes(x=x1)) + geom\_histogram(**binwidth**=0.1)
  + ggplot(df, aes(x=x1)) + geom\_histogram(aes(y=**..density..**), binwidth=0.1)
  + ggplot(df, aes(x=x1, fill=cat\_var)) + geom\_histogram(binwidth=0.1, **position**=”stack/fill/dodge”)

|  |  |  |
| --- | --- | --- |
|  |  |  |

* **Bar Plot**: geom\_bar() : x-axis = categorical variables
  + ggplot(df, aes(x=cat\_var)) + geom\_bar(stat=”bin”)
  + Custom Color Palettes

blues <- brewer.pal(9, "Blues")

blue\_range <- colorRampPalette(blues)

ggplot(Vocab, aes(x = education, fill = vocabulary)) +

geom\_bar(position = "fill") +

scale\_**fill\_manual**(**values**=blue\_range(11))

* + Overlapping bar plots

posn\_d <- **position\_dodge**(**width**=0.2)

ggplot(mtcars, aes(x = cyl, fill = am)) + geom\_bar(position=posn\_d)

* **Line Plots:** geom\_line()
  + Plotting different categories

ggplot(df, aes(x=Year, y=Capture, **linetype**=Species)) + geom\_line()

* + Proportional Trends

ggplot(df, aes(x=Year, y=Capture, **fill**=Species)) + geom\_**area**(position=”fill”)

* + ggplot(economics, aes(x=date, y=unemploy/pop)) + **geom\_rect**(data=recess, aes(xmin=begin, xmax=end, ymin=-Inf, ymax=+Inf), **inherit.aes=FALSE**, fill="red", alpha=0.2) + geom\_line()

**qplot**

* Quick and dirty way for plotting, not very flexible, doesn’t follow grammar of graphics
* qqplot( x, y, data, shape/size/col, postion, jitter, alpha=I(value) )

**Wrap-Up**

