

Agenda

General

- Announcement: Homework plan
- Grades: on Brightspace this week
- Grades: must submit before class for your submission to 'count'

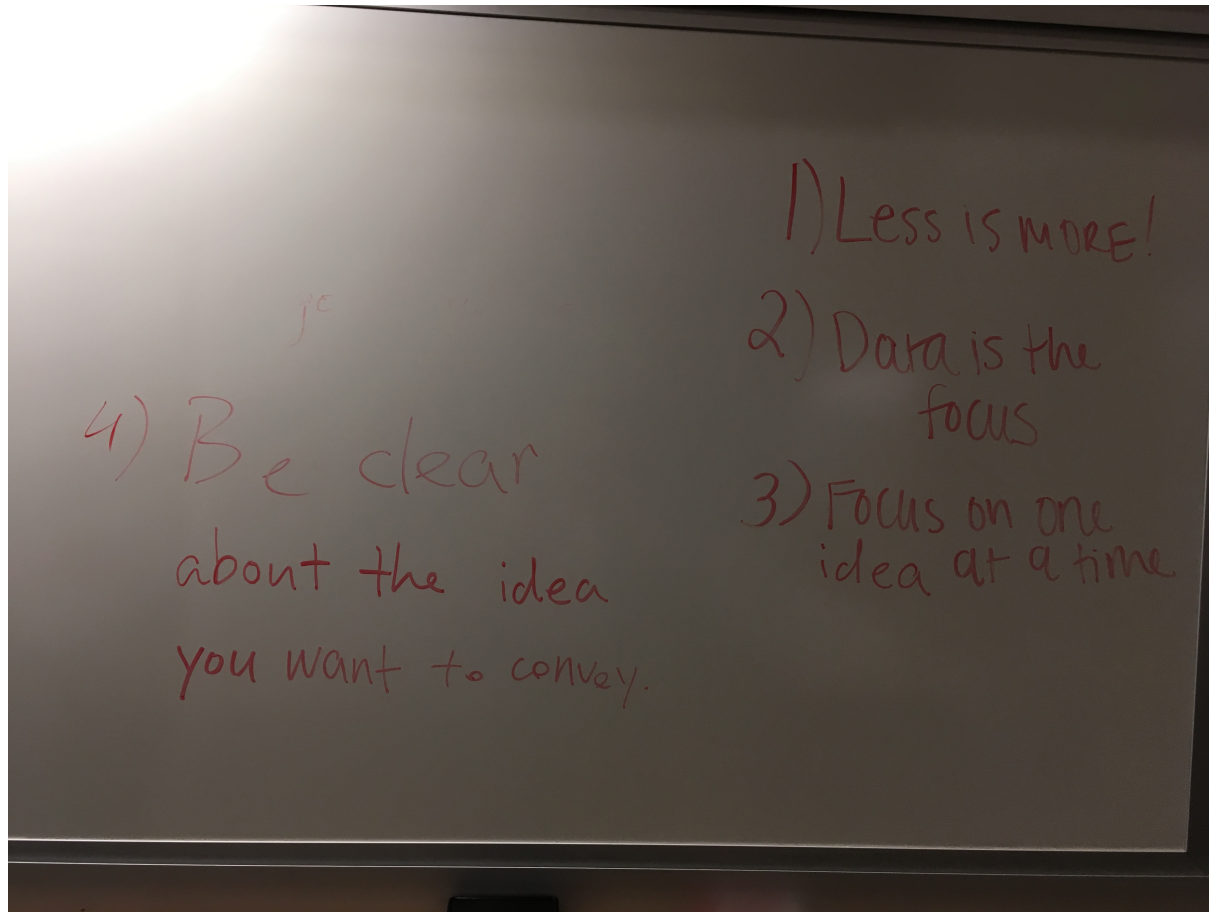
Class Focus

- short review
- grammar of graphics!
- ggplot introduction
- Scatterplots & bivariate data

What did we learn last week?

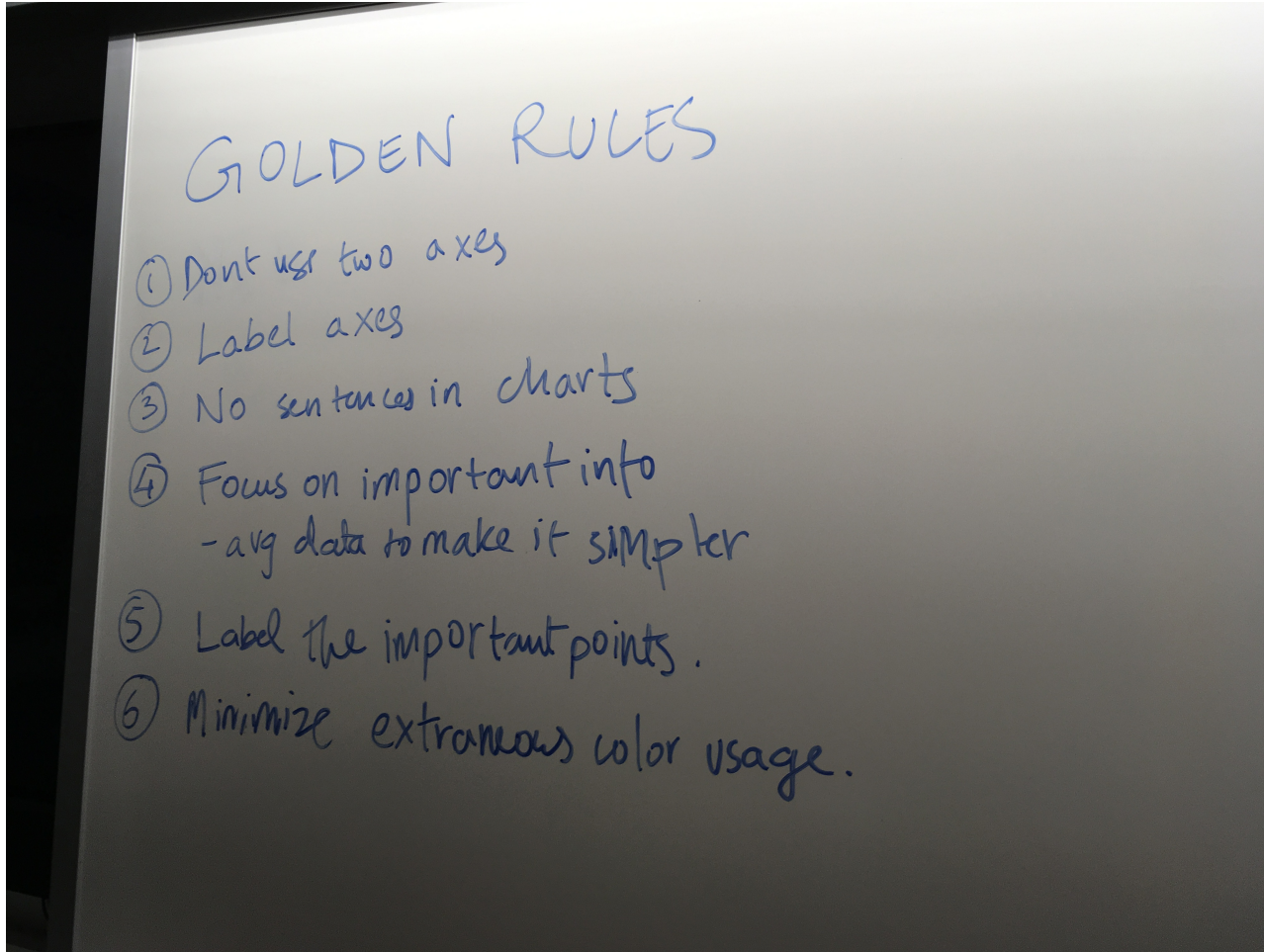
- How to use `data.table` for quick assessments of the data
- How this relates to tidy data
- The difference between wide and long data
- How to understand what makes a graphic effective

Golden Rules



- 1) less is more!
- 2) data is the focus
- 3) focus on one idea at a time
- 4) be clear about the idea you want to convey

Golden Rules



- 1) Don't use two axes
- 2) Label axes
- 3) No sentences in charts
- 4) Focus on important info / simplify
- 5) Minimize extraneous color usage

Golden Rules

- Pick the **R**ight type of graph
- Relevant **dA**ta only
- Consistent **S**cale/appropriate axis
- Color & **conT**rast for emphasis
- if not important, **K**ick it out
- have **lO**ts of fun

- 1) pick the Right type of graph
- 2) relevant data only
- 3) consistent scale /appropriate axes
- 4) color and contrast for emphasis
- 5) if not important, kick it out
- 6) have lots of fun

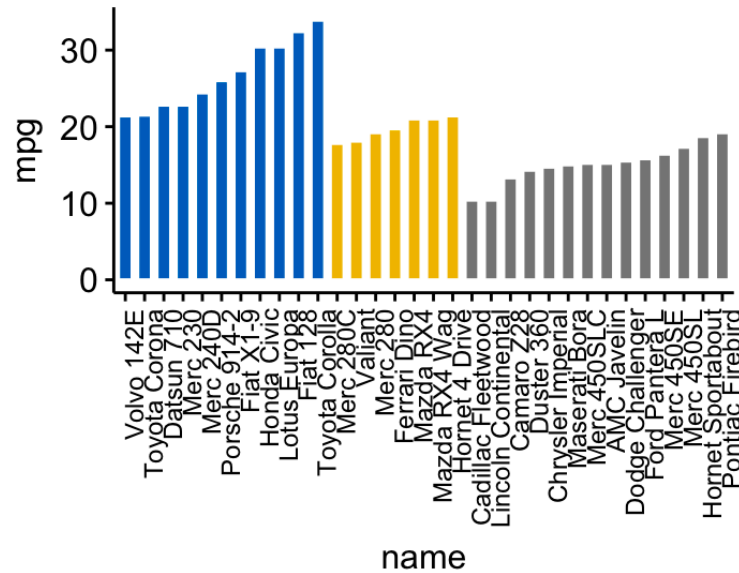
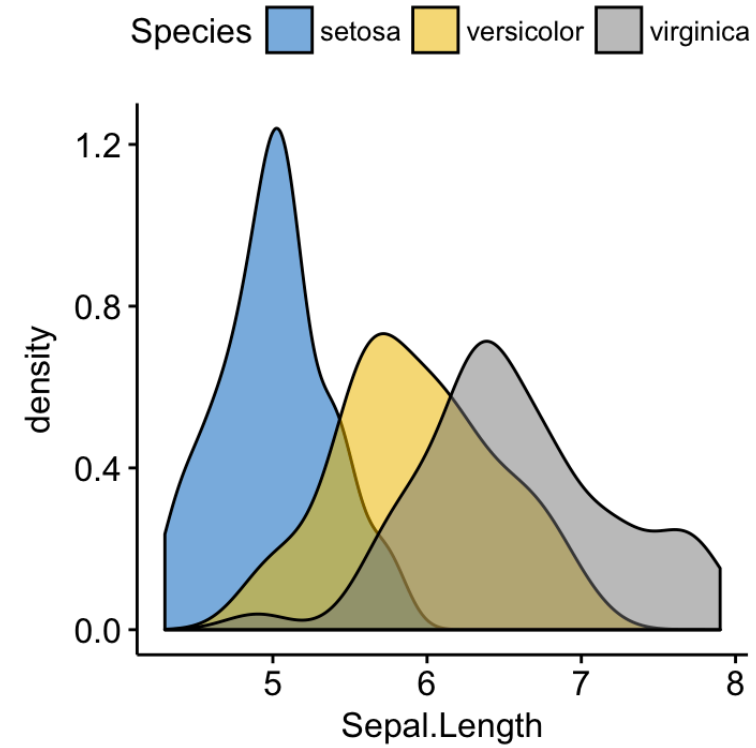
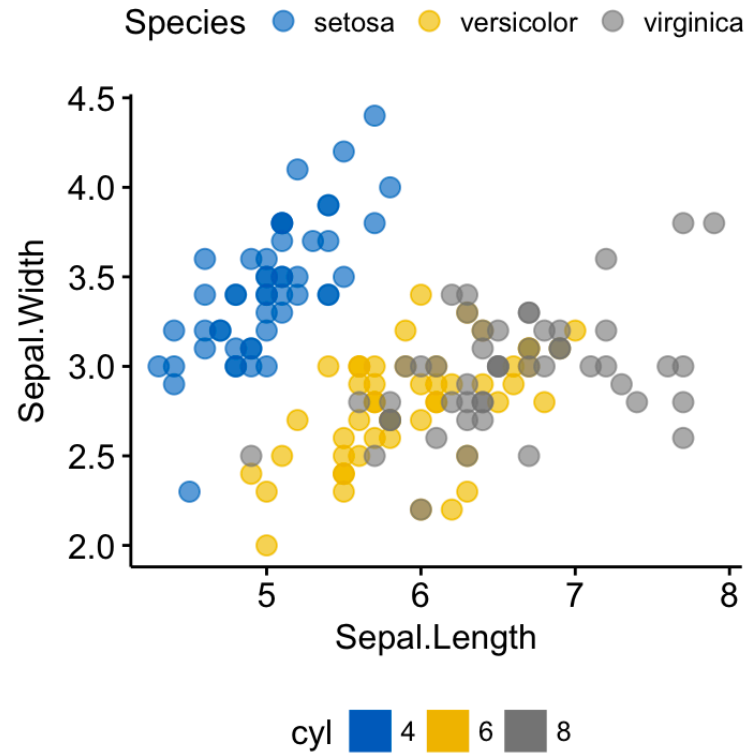
Golden Rules

1. One y-axis
2. No text in figure body
3. Embrace negative space
4. Keep message simple
5. Reduce dimensions
6. Be honest with scales

- 1) one y-axis
- 2) no text in figure body
- 3) embrace negative space
- 4) keep message simple
- 5) reduce dimensions
- 6) be honest with scales

Introduction to ggplot

Week 3, day 1
EDA
Dr. Dorff

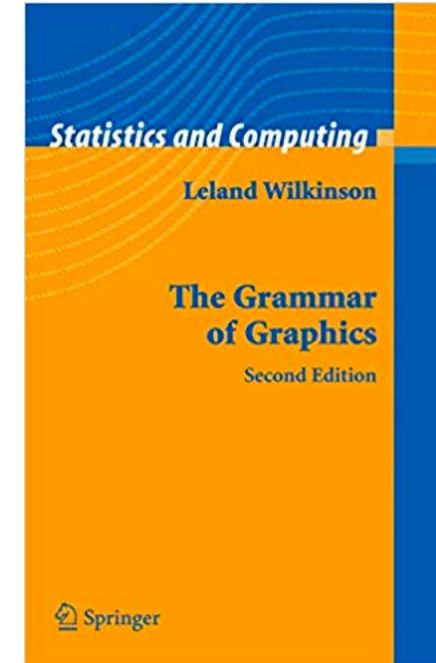


Species	length	mean	sd
setosa	50	5.01	0.352
versicolor	50	5.94	0.516
virginica	50	6.59	0.636

iris data set gives the measurements in cm of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

Wilkinson's Grammar of Graphics

- A good **grammar** will allow us to gain insight into the **composition** of complicated graphics, and reveal unexpected connections between seemingly different graphics



What Is The Grammar Of Graphics?

- Building blocks of a graph include:
 - data
 - aesthetic mapping
 - geometric object
 - statistical transformations
 - scales
 - coordinate system
 - position adjustments
 - faceting

(1) START WITH THE DATA

Table 1. Simple dataset.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b




Mapped to aesthetics
for a graph

<i>x</i>	<i>y</i>	Shape
2	4	a
1	1	a
4	15	b
9	80	b

(2) SCALES & SHAPES

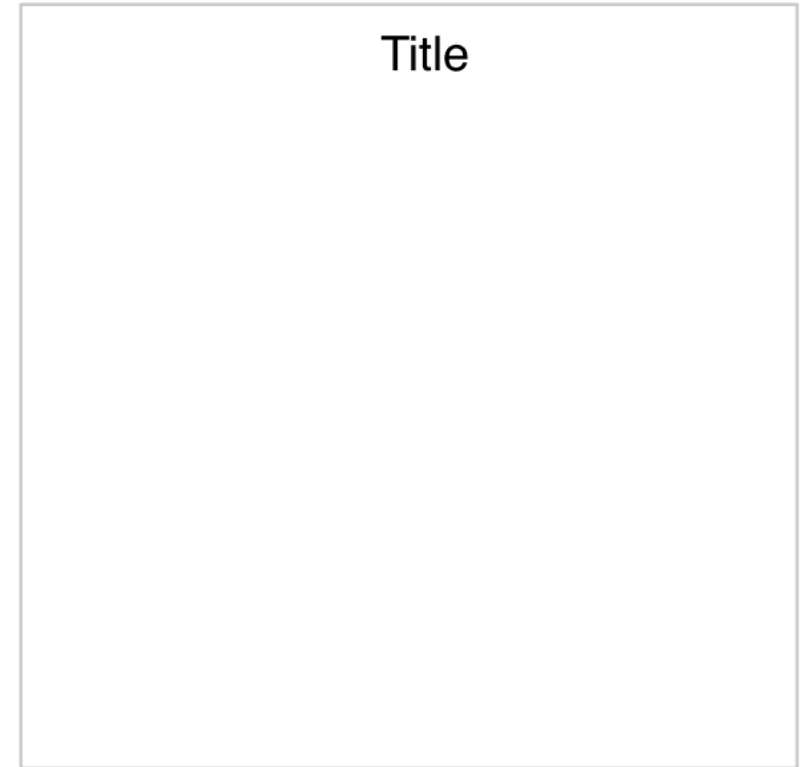
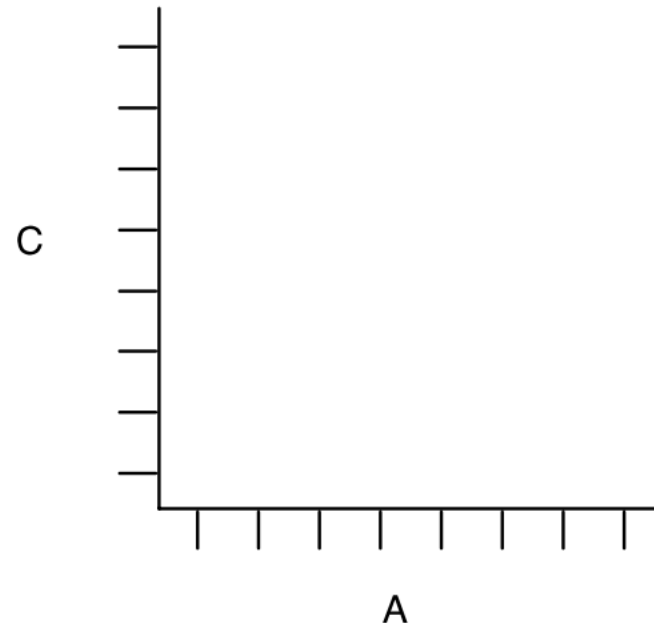
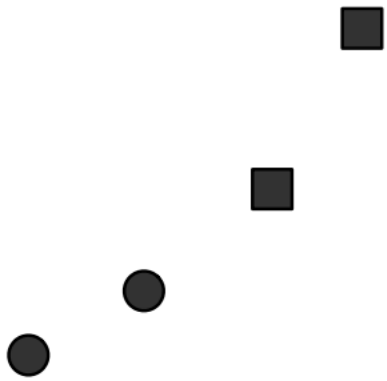
x	y	Shape
2	4	a
1	1	a
4	15	b
9	80	b

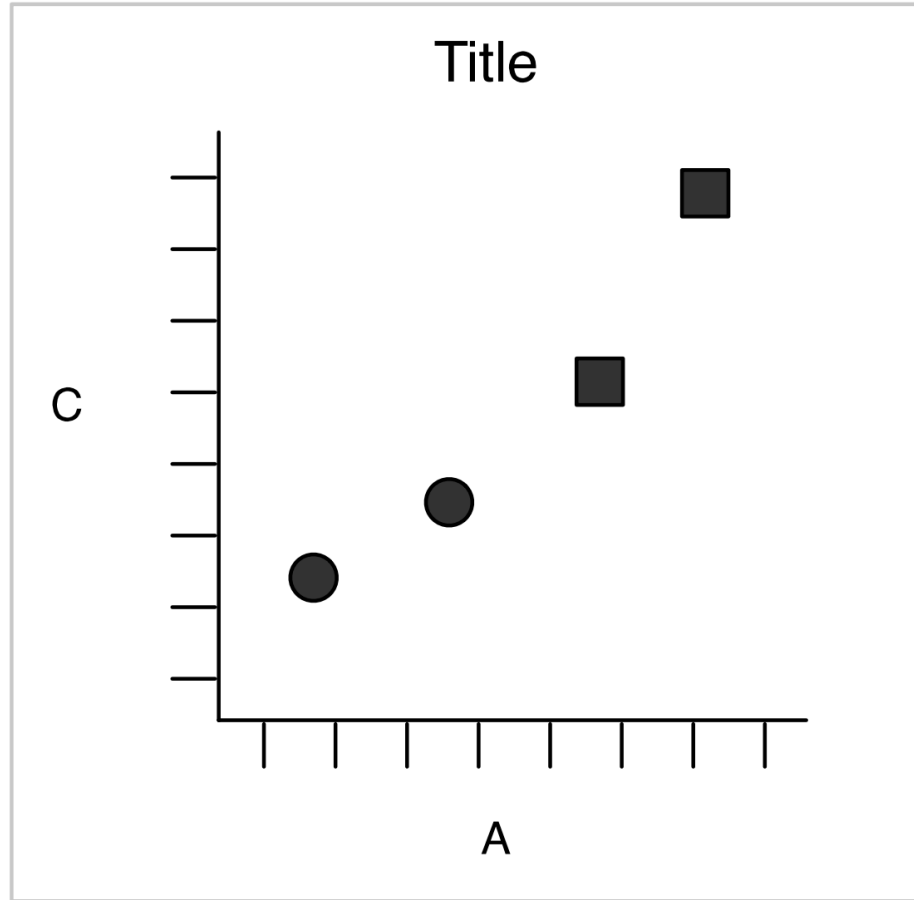


x	y	Shape
25	11	circle
0	0	circle
75	53	square
200	300	square

variables mapped into aesthetic space.

geometric objects, scales and coordinate system, plot annotations.





layered grammar (revisited)

- a default dataset and set of mappings from variables to aesthetics
- one or more layers, with each layer having one geometric object
- one statistical transformation
- one scale for each aesthetic mapping used
- a coordinate system
- the facet specification.

DATA

MAPPING

```
ggplot(diamonds, aes(carat, price))  
+ geom_point()
```

The diagram shows the R code for creating a ggplot. The word 'DATA' is written in purple above a bracket that spans the 'diamonds' argument of the 'ggplot' function. The word 'MAPPING' is written in purple above an arrow that points to the 'aes' function. The 'aes' function is highlighted in yellow. A bracket above 'carat' is marked with an 'x', and a bracket above 'price' is marked with a 'y', indicating their roles as x and y coordinates in the plot.

THESE ARE LAYERS!