

Graphics in R: ggplot2

a **statsTeachR** resource

These slides were created for **statsTeachR** by Nicholas Reich and are released under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

Let there be data....

```
library(NHANES)
str(NHANES)

## 'data.frame': 31126 obs. of 32 variables:
## $ seqn      : num  1 2 3 4 5 6 7 8 9 10 ...
## $ sex       : Factor w/ 2 levels "male","female": 2 1 2 1 1 2 2 1 2 1 ...
## $ age        : num  2 77 10 1 49 19 59 13 11 43 ...
## $ pregnant   : Factor w/ 2 levels "yes","no": 2 2 2 2 2 2 2 2 2 2 ...
## $ ethnicity  : Factor w/ 5 levels "Mexican American",...: 4 3 3 4 3 5 4 3 4 4 ...
## $ death      : Factor w/ 3 levels "alive","other",...: NA 1 NA NA 1 1 1 NA NA ...
## $ followup   : num  NA 90 NA NA 74 86 76 NA NA 79 ...
## $ smoker     : Factor w/ 2 levels "yes","no": 2 2 2 2 1 2 2 2 2 2 ...
## $ diabetic   : num  0 0 0 0 0 0 0 0 0 0 ...
## $ height     : num  0.916 1.74 1.366 NA 1.783 ...
## $ weight     : num  12.5 75.4 32.9 13.3 92.5 ...
## $ waist      : num  0.457 0.98 0.647 NA 0.999 0.816 0.907 0.641 0.646 1.08 ...
## $ wci        : num  0.0789 0.0871 0.0817 NA 0.0791 ...
## $ bmi        : num  14.9 24.9 17.6 NA 29.1 ...
## $ ptfat      : num  NA 14.3 NA NA 16.5 ...
## $ tfat        : num  NA 10.8 NA NA 15.2 ...
## $ lfat        : num  NA 11.4 NA NA 12.7 ...
## $ llean       : num  NA 22 NA NA 28.6 ...
## $ lbmi        : num  NA 162935 NA NA 194415 ...
## $ fbmi        : num  NA 77162 NA NA 91899 ...
## $ bbmi        : num  NA 9009 NA NA 8134 ...
```

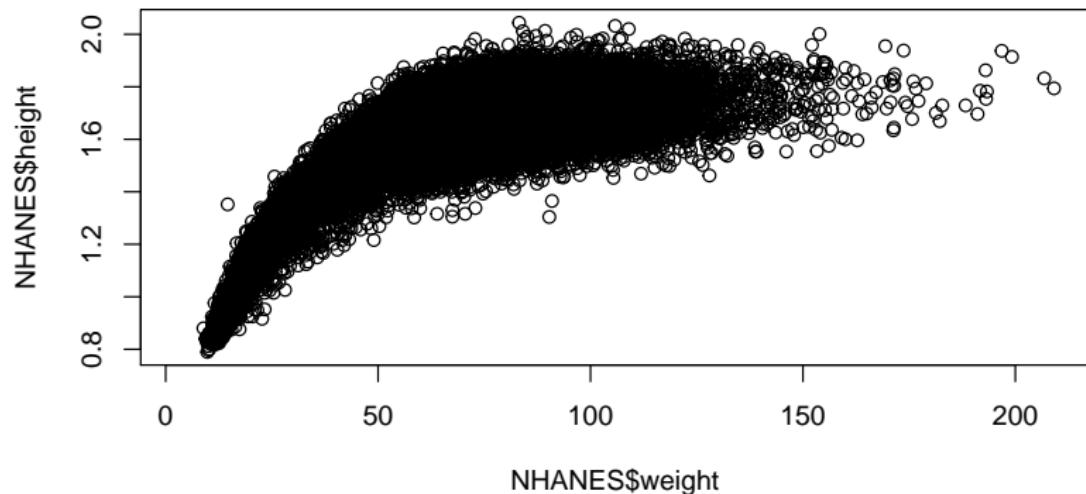
Graphics in R

You have three central choices for making graphics in R

- ▶ “Base graphics”
- ▶ ggplot2
- ▶ lattice

Graphics in R: Base graphics

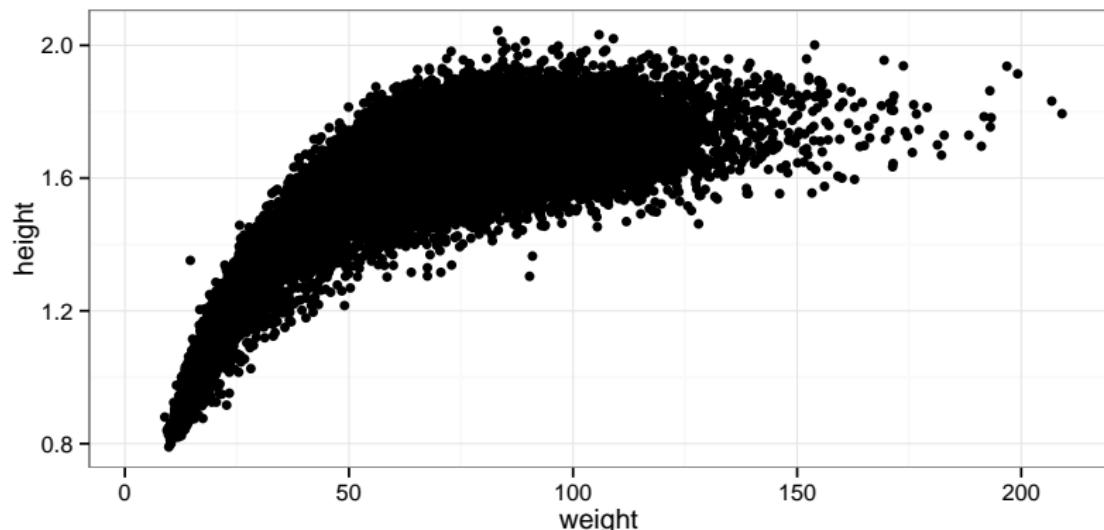
```
plot(NHANES$weight, NHANES$height)
```



A simple ggplot2 graphic

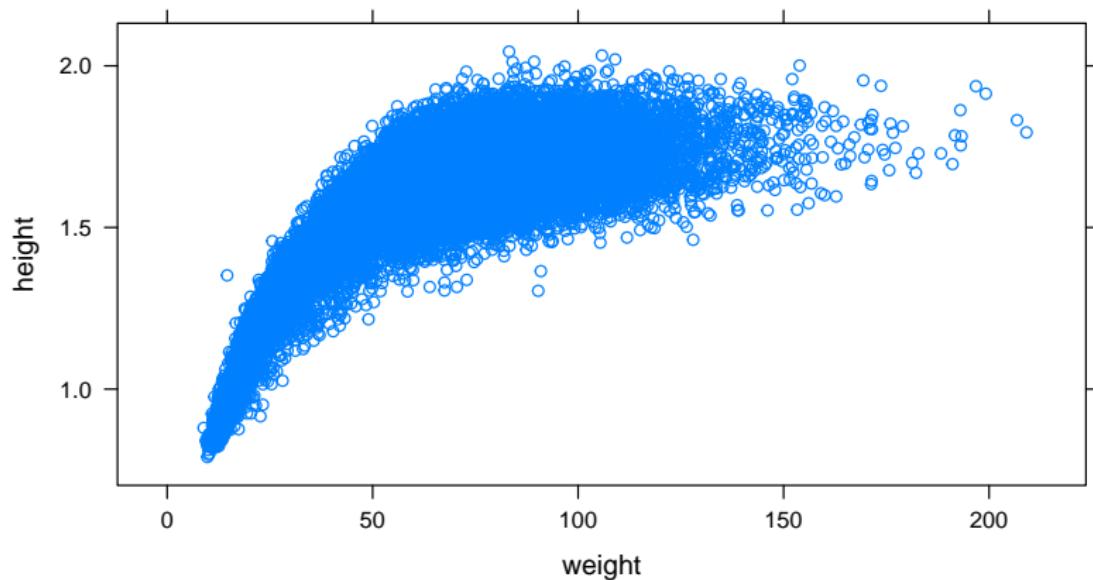
```
library(ggplot2)
qplot(weight, height, data=NHANES)

## Warning: Removed 4967 rows containing missing
values (geom_point).
```



A simple lattice graphic

```
library(lattice)
xyplot(height ~ weight, data=NHANES)
```



The mosaic package can translate between them

```
library(mosaic)  
mplot(NHANES)
```

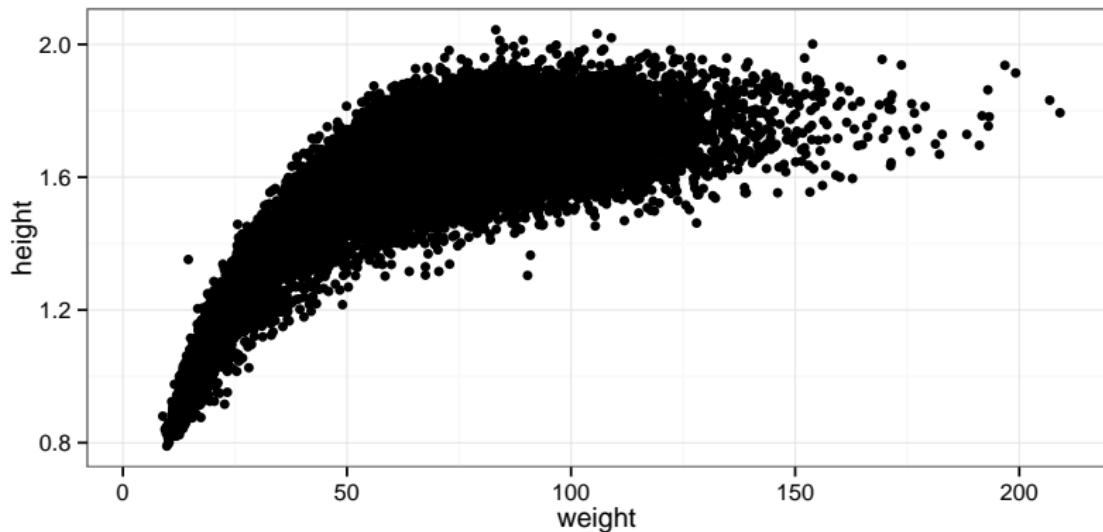
`mplot()` loads a very nice graphical user interface for creating graphics, and spits out the code for `ggplot` or `lattice` graphics.
Try it out! Practice saving a graphic from RStudio.

Why I use ggplot2

- ▶ “easy” to create both simple and complex graphics
- ▶ unified syntax
- ▶ attractive defaults (except for colors)
- ▶ actively developed and improved

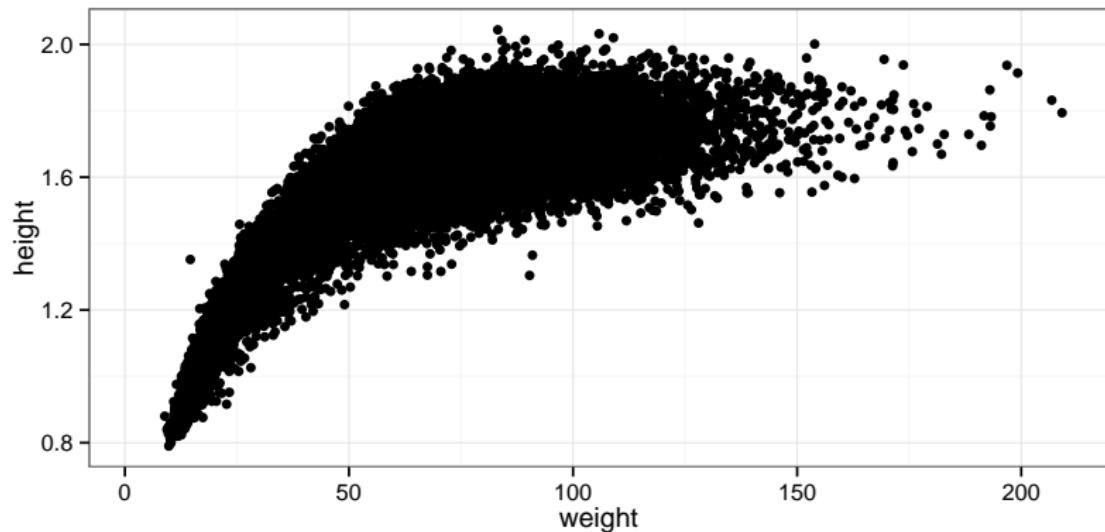
Quick tour of ggplot2 syntax

```
## defaults to scatterplot
qplot(weight, height, data=NHANES)
```



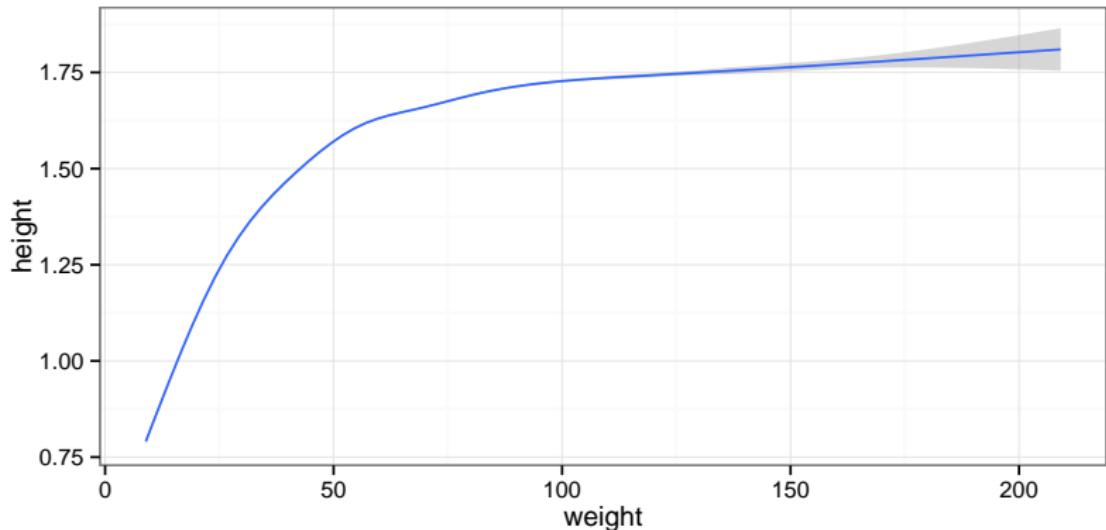
Quick tour of ggplot2 syntax

```
## same as default  
qplot(weight, height, data=NHANES, geom="point")
```



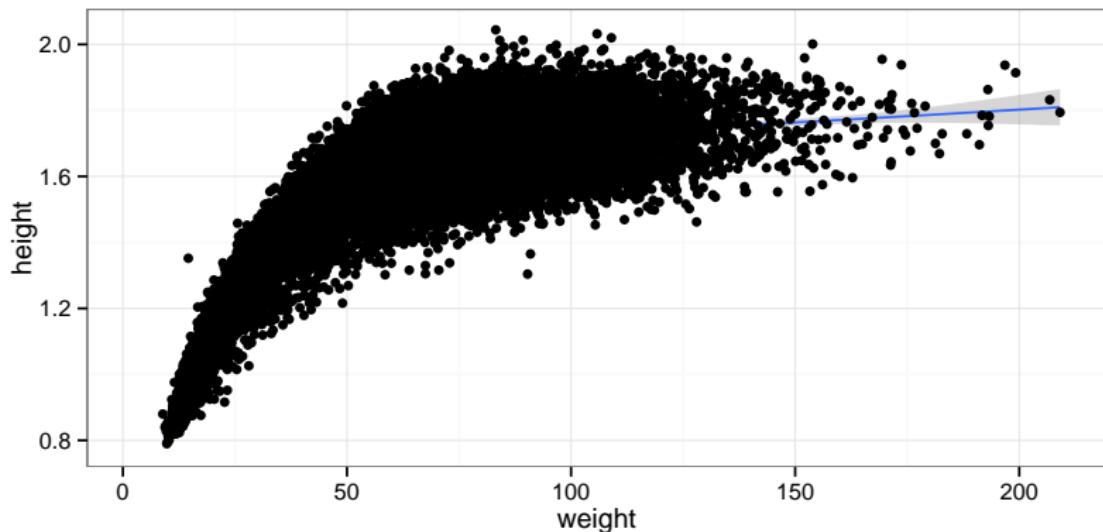
Quick tour of ggplot2 syntax

```
qplot(weight, height, data=NHANES, geom="smooth")
```



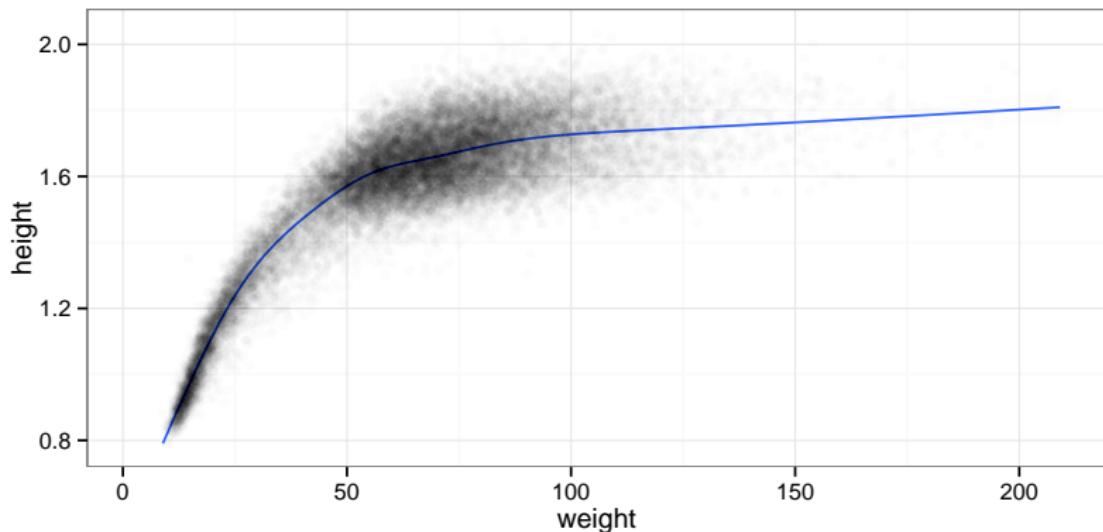
Quick tour of ggplot2 syntax

```
qplot(weight, height, data=NHANES,  
      geom=c("smooth", "point"))
```



Quick tour of ggplot2 syntax

```
qplot(weight, height, data=NHANES,  
      geom=c("smooth", "point"), alpha=I(.01))
```



ggplot2 building blocks

The grammar ...

- ▶ geom
- ▶ aesthetics (aes)
- ▶ scales
- ▶ facets
- ▶ data
- ▶ ... and more here: <http://docs.ggplot2.org/current/>

What is a “geom”?

From Hadley:

- ▶ Geoms define the basic “shape” of the elements on the plot
- ▶ Basics: point, line, bar, text, hline, vline
- ▶ Statistics: histogram, smooth, density
- ▶ Others: boxplot, pointrange, linerange, ribbon

For more info check out the documentation:

<http://docs.ggplot2.org/current/>.

What are “aesthetics”?

Aesthetics define a mapping between data and the display.

length	width	depth	trt
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b



x	y	colour
2	3	a
1	2	a
4	5	b
9	10	b

Thinking about aesthetics

Each geom has a different set of aesthetics.

What aesthetics might we need for geom_point?

Aesthetics for geom_point

`geom_point` understands the following aesthetics

- ▶ `x` (required)
- ▶ `y` (required)
- ▶ `alpha`
- ▶ `color`
- ▶ `fill`
- ▶ `shape`
- ▶ `size`

Aesthetics for geom_line

What aesthetics might we need for geom_line?

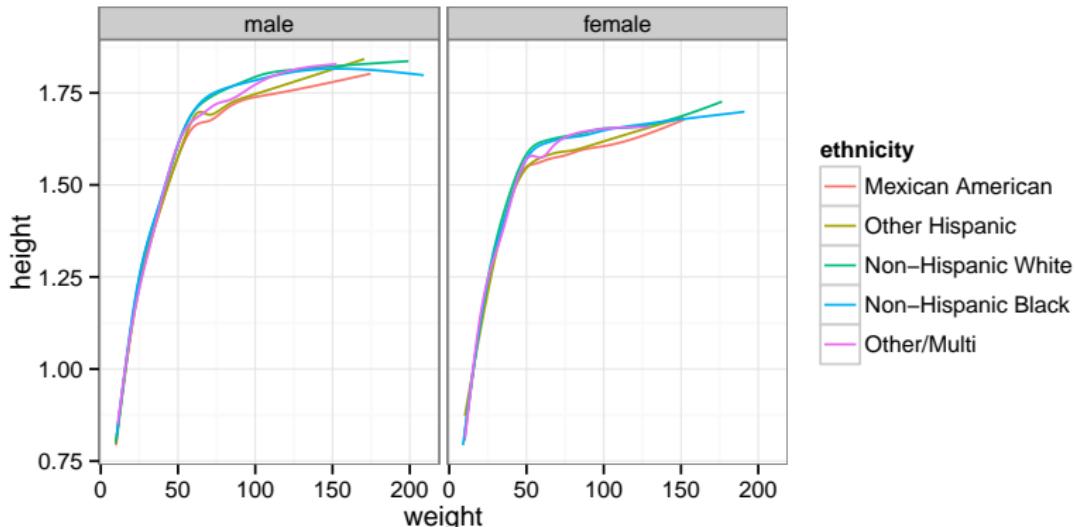
Aesthetics for geom_line

What aesthetics might we need for geom_line?

- ▶ x (required)
- ▶ y (required)
- ▶ alpha
- ▶ color
- ▶ linetype
- ▶ size

A more involved example

```
qplot(weight, height, data=NHANES,  
      geom="smooth", se=FALSE,  
      color=ethnicity,  
      facets=.~sex)
```



qplot() vs ggplot

```
qplot(weight, height, data=NHANES,  
      geom="smooth", se=FALSE,  
      color=ethnicity,  
      facets=.~sex)
```

```
ggplot(NHANES) +  
  geom_smooth(aes(x=weight, y=height, color=ethnicity), se=FALSE) +  
  geom_point(aes(x=weight, y=height, color=ethnicity), alpha=I(.01)) +  
  facet_grid(.~sex)
```

How to get help with ggplot

- ▶ Ask Google: "ggplot [what I'm trying to do]"
- ▶ Ask your classmates/instructors on Piazza!
- ▶ Post something on stackexchange or the ggplot2 mailing list
- ▶ Read more!
 - ▶ <http://ggplot2.org/>
 - ▶ Cookbook for R
 - ▶ Hadley's book
- ▶ Practice, practice, practice.