Statistically Speaking Lecture Series

Sponsored by the Biostatistics Collaboration Center

Using R for Statistical Graphics: The Do's and Don'ts of Data Visualization

David Aaby, MS

Senior Statistical Analyst, Preventive Medicine

Northwestern Medicine®
Feinberg School of Medicine

BCC: Biostatistics Collaboration Center

Who We Are

Leah J. Welty, PhD
Assoc. Professor
BCC Director

Lauren Balmert, PhD
Asst. Professor

Jody D. Ciolino, PhD
Asst. Professor

Kwang-Youn A. Kim, PhD
Assoc. Professor

Masha Kocherginsky, PhD
Assoc. Professor

Mary J. Kwasny, ScD
Assoc. Professor

Julia Lee, PhD, MPH
Assoc. Professor

David Aaby, MS
Senior Stat. Analyst

Elizabeth Gray, MS
Stat. Analyst

Kimberly Koloms, MS
Stat. Analyst

Amy Yang, MS
Senior Stat. Analyst

Tameka L. Brannon
Financial | Research
Administrator

Biostatistics Collaboration Center (BCC)

Mission: to support investigators in the conduct of high-quality, innovative health-related research by providing expertise in biostatistics, statistical programming, and data management.

How do we accomplish this?

1. Every investigator is provided a **FREE** initial consultation of 1-2 hours, subsidized by **FSM Office for Research**. Thereafter:
   a) Grants
   b) Subscription
   c) Re-charge (Hourly) Rates

2. Grant writing (e.g. developing analysis plans, power/sample size calculations) is also supported by FSM at **no cost to the investigator**, with the goal of establishing successful collaborations.

BCC: Biostatistics Collaboration Center

What We Do

- Many areas of expertise, including:
  - Bayesian Methods
  - Big Data
  - Bioinformatics
  - Causal Inference
  - Clinical Trials
  - Database Design
  - Genomics
  - Longitudinal Data
  - Missing Data
  - Reproducibility
  - Survival Analysis

Many types of software, including:

BCC: Biostatistics Collaboration Center

Shared Statistical Resources

Northwestern Medicine | NUCATS Clinical and Translational Sciences Institute

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Stanley Manne Children's Research Institute | Ann & Robert H. Lurie Children's Hospital of Chicago

Biostatistics Collaboration Center (BCC)

- Supports non-cancer research at NU
- Provides investigators an initial 1-2 hour consultation subsidized by the FSM Office of Research
- Grant, Hourly, Subscription

Shirley Ryan AbilityLab

Quantitative Data Sciences Core (QDSC)

- Supports all cancer-related research at NU
- Provides free support to all Cancer Center members subsidized by RHLCCC
- Grant

Biostatistics Research Core (BRC)

- Supports Lurie Children's Hospital affiliates
- Provides investigators statistical support subsidized by the Stanley Manne Research Institute at Lurie Children's.
- Hourly

BCC: Biostatistics Collaboration Center

Shared Resources Contact Info

- Biostatistics Collaboration Center (BCC)
  - Website: http://www.feinberg.northwestern.edu/sites/bcc/index.html
  - Email: bcc@northwestern.edu
  - Phone: 312.503.2288

- Quantitative Data Sciences Core (QDSC)
  - Website: http://cancer.northwestern.edu/research/shared_resources/quantitative_data_sciences/index.cfm
  - Email: qdsc_rhlccc@northwestern.edu
  - Phone: 312.503.2288

- Biostatistics Research Core (BRC)
  - Website: https://www.luriechildrens.org/en-us/research/facilities/Pages/biostatistics.aspx
  - Email: mereed@luriechildrens.org
  - Phone: 773.755.6328

Northwestern Medicine®
Feinberg School of Medicine

Using R for Statistical Graphics: The Do's and Don'ts of Data Visualization

Outline

1. A (good) picture is worth 1,000 words: A motivating example
2. Data Visualization: The Good, The Bad, and The Ugly
3. Why should I use R?
4. Using R for Creating Statistical Graphics: A brief walkthrough

Northwestern Medicine®
Feinberg School of Medicine

"A (good) picture is worth 1,000 words"

Good Pictures: An example of why they're needed

Correlation Defined

Correlation often denoted as r

Measures strength of the *linear* association between two continuous variables

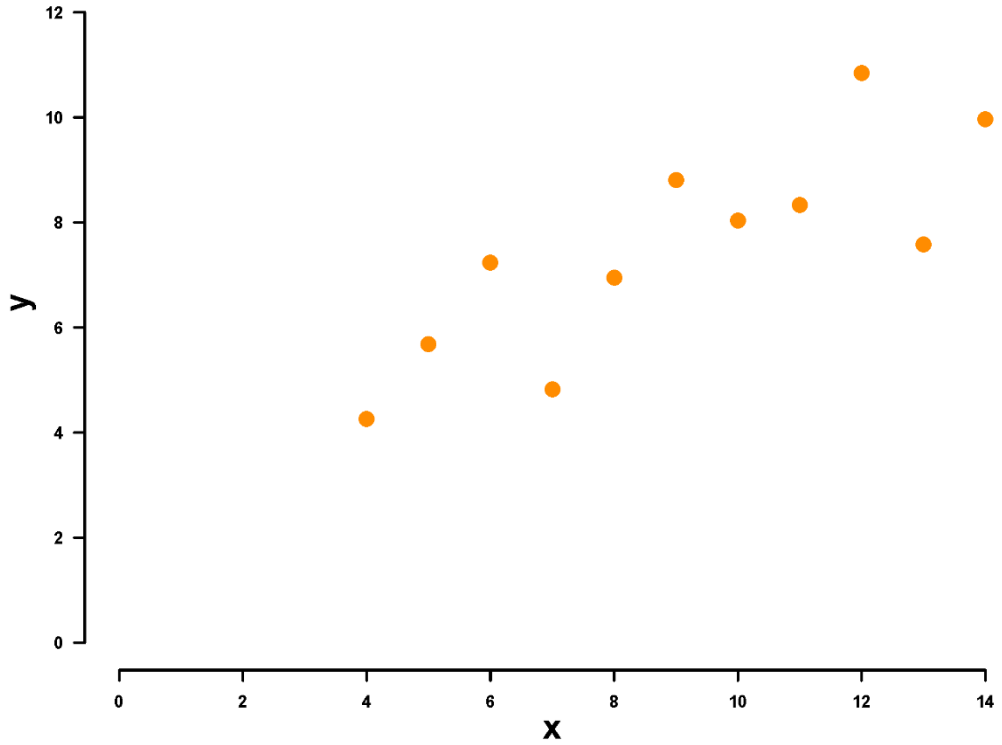
$$-1 \leq r \leq 1$$

r = -1	strong negative linear association
r = 1	strong positive linear association
r = 0	no linear association

Good Pictures: An example of why they're needed

Correlation Example

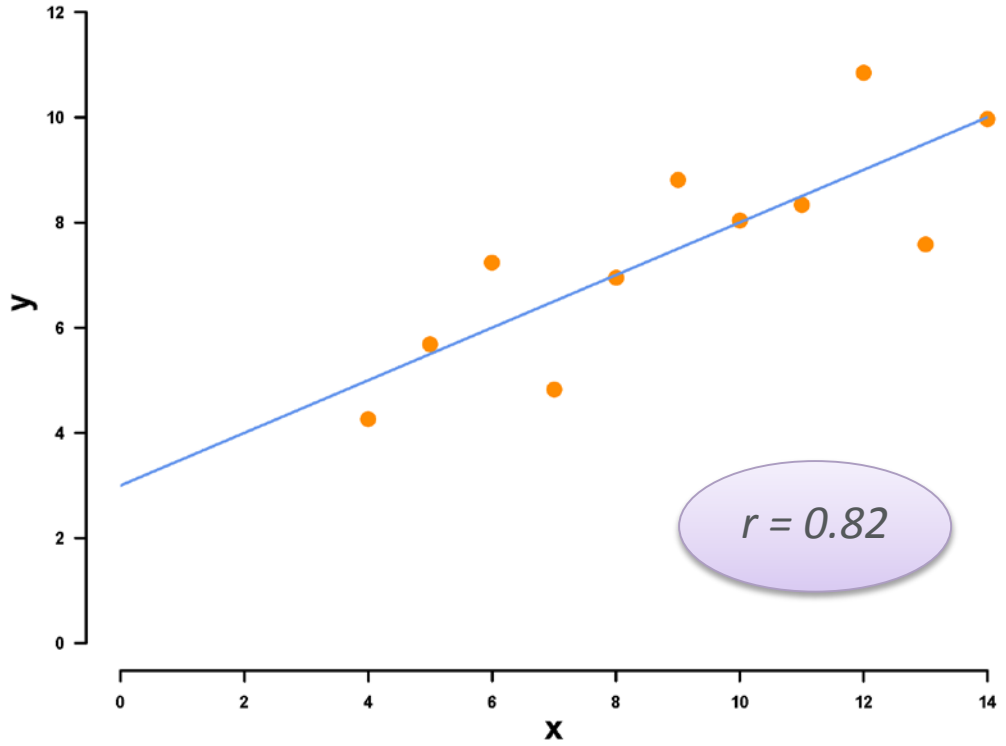
- Is there a strong positive linear association between x and y?
- The correlation between two variables of interest, x and y, is 0.82



Good Pictures: An example of why they're needed

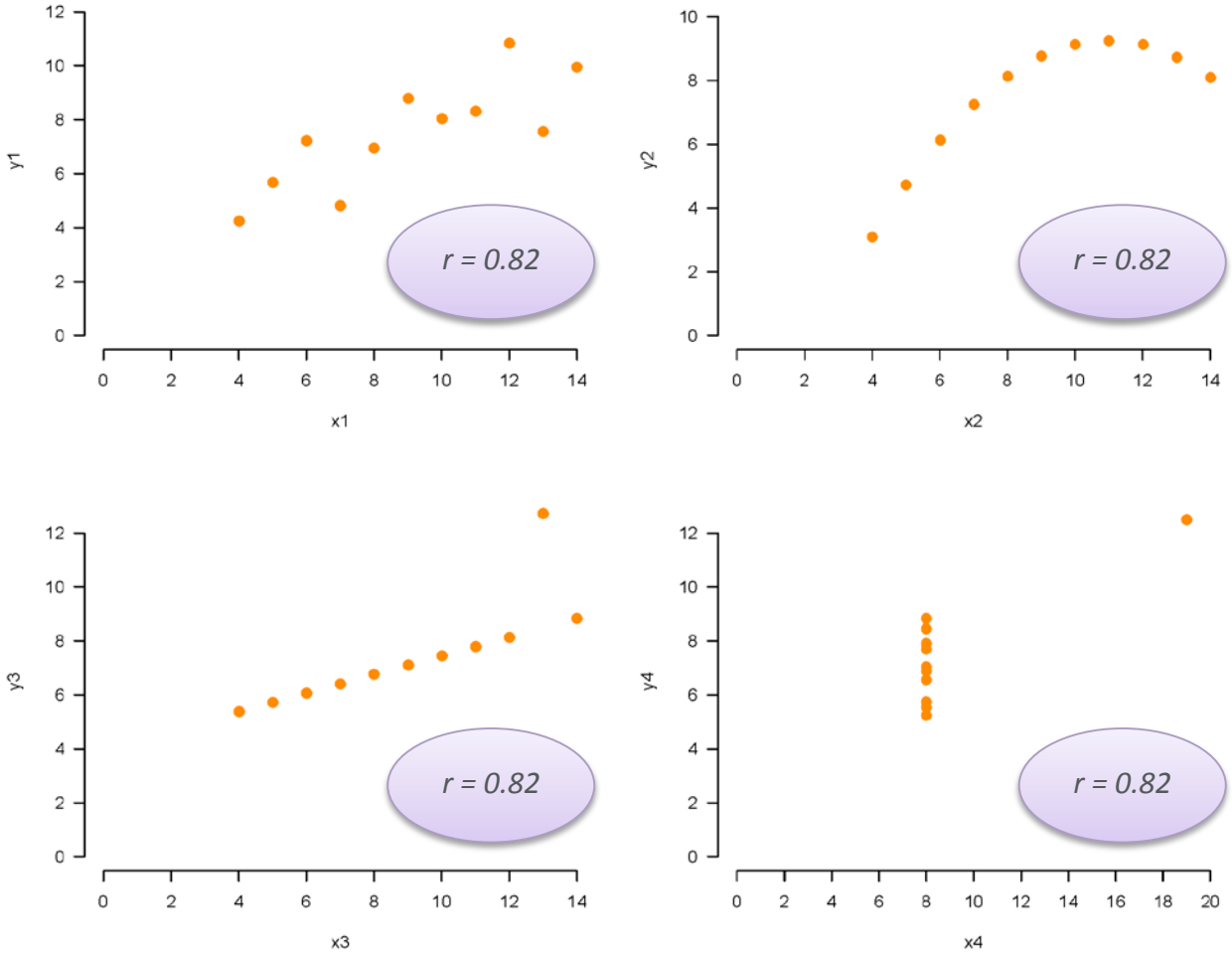
Correlation Example

- Is there a strong positive linear association between x and y?
- The correlation between two variables of interest, x and y, is 0.82



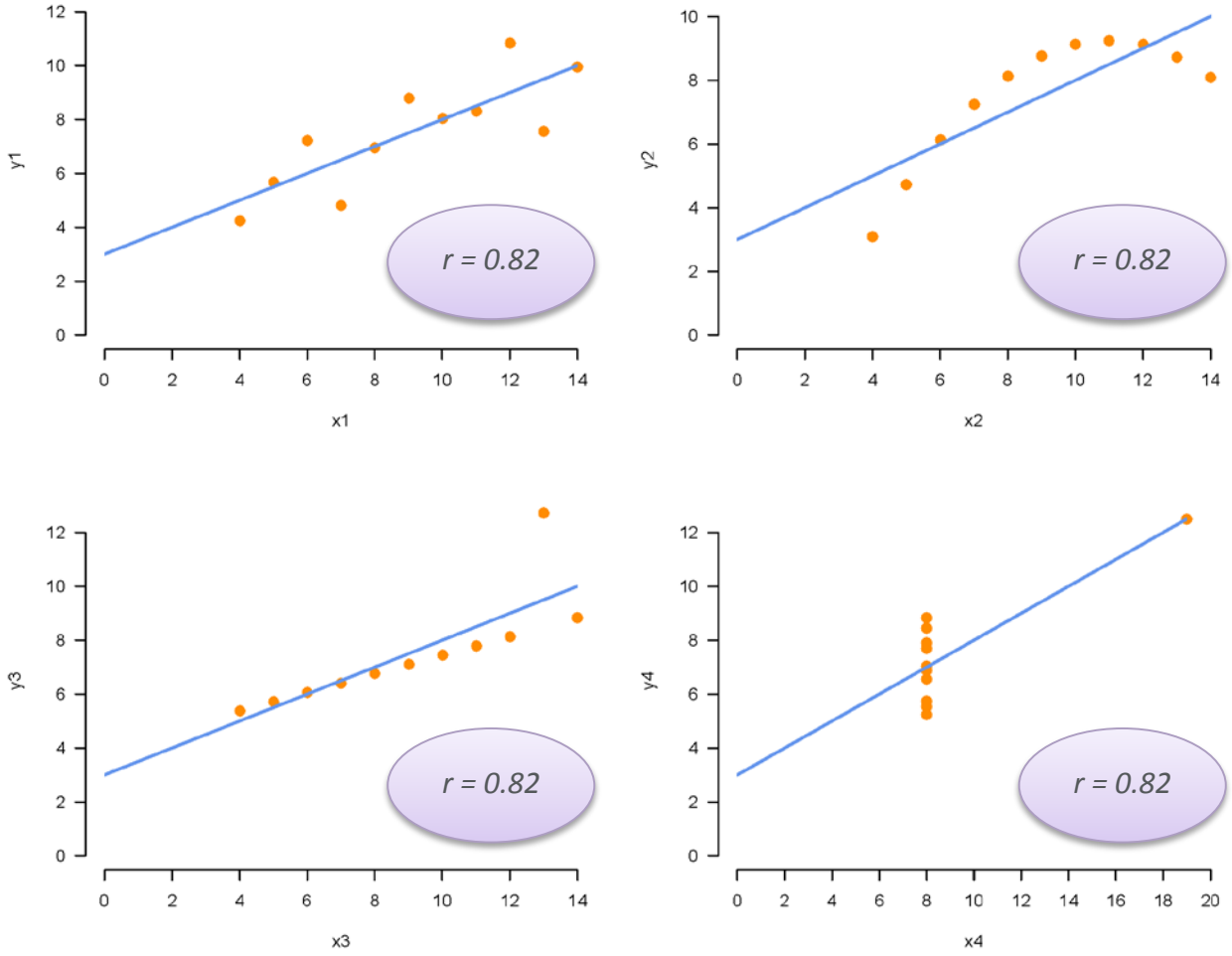
Good Pictures: An example of why they're needed

Anscombe's Quartet



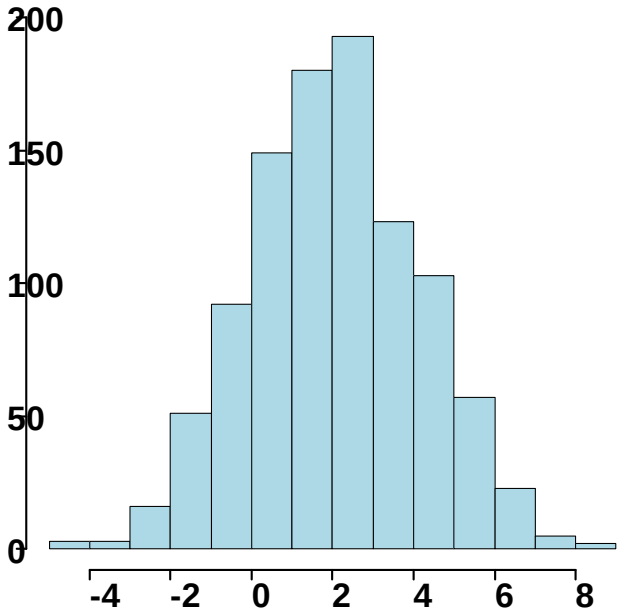
Good Pictures: An example of why they're needed

Anscombe's Quartet

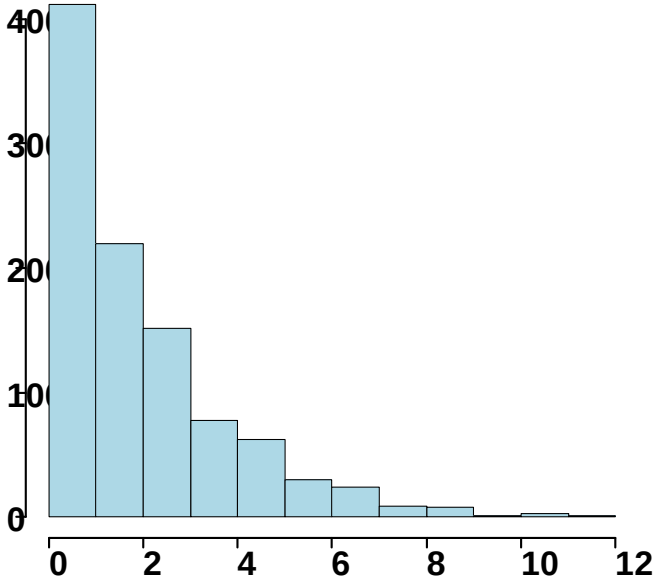


Above Average: Picture Your Data!

What do you think of when you hear “The mean value was 2.0”?



What we tend to think
Mean = 2
Median = 2



What might be true
Mean = 2.0
Median = 1.4

Northwestern Medicine®
Feinberg School of Medicine

Data Visualization

The Good, The Bad, and The Ugly

Goals of data visualization

What makes a good graphic?

- Communicate information clearly and efficiently
- Summarize data, help inform analytic techniques
- Illustrate trends, patterns, relationships in data not otherwise seen in a table of values or measurements
- Make results easy to digest and understand

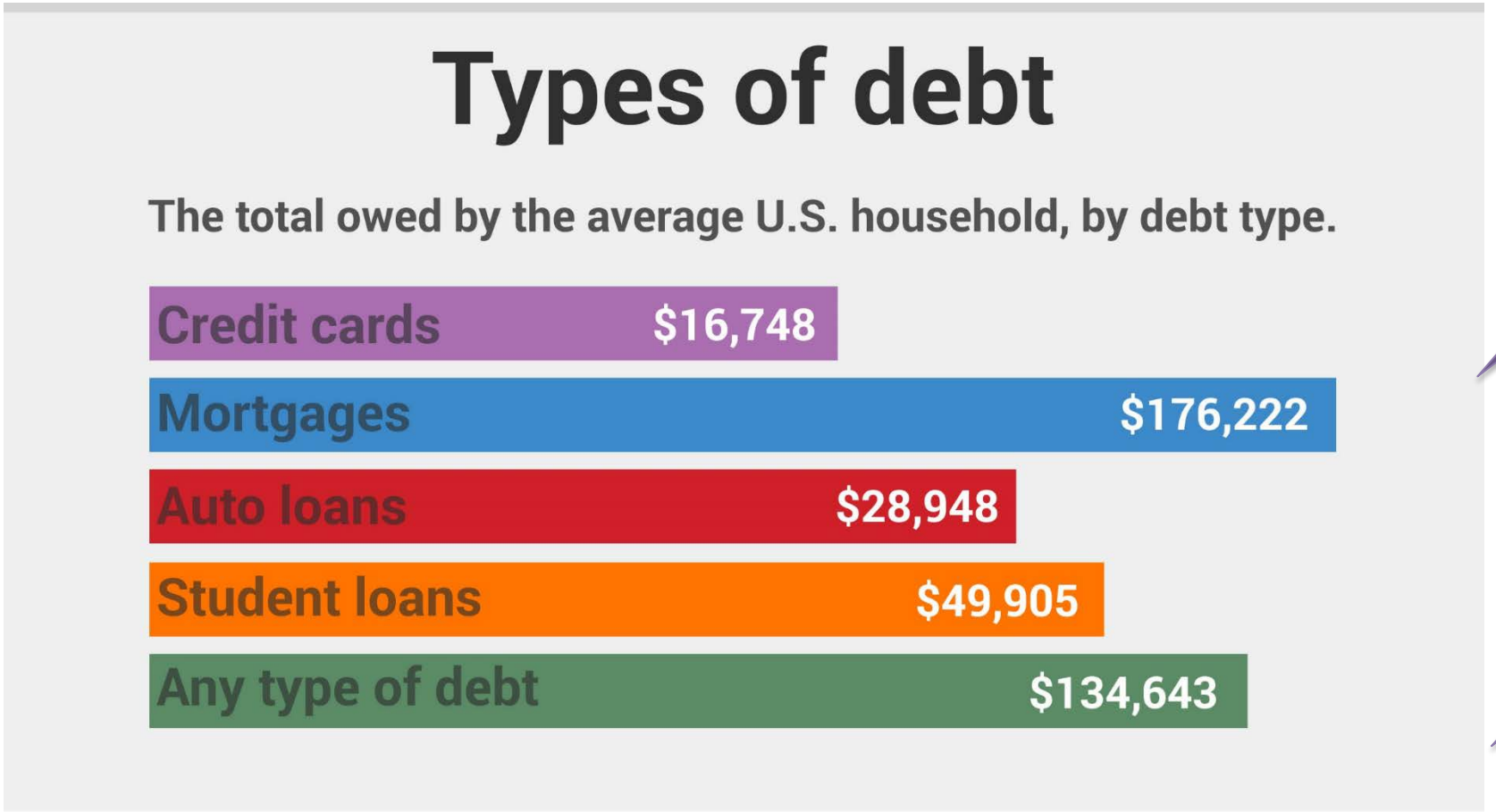
Let’s look at some examples

Can you tell which graphs are good, bad, or ugly?

https://frinkiac.com/

Example 1

Bad graph

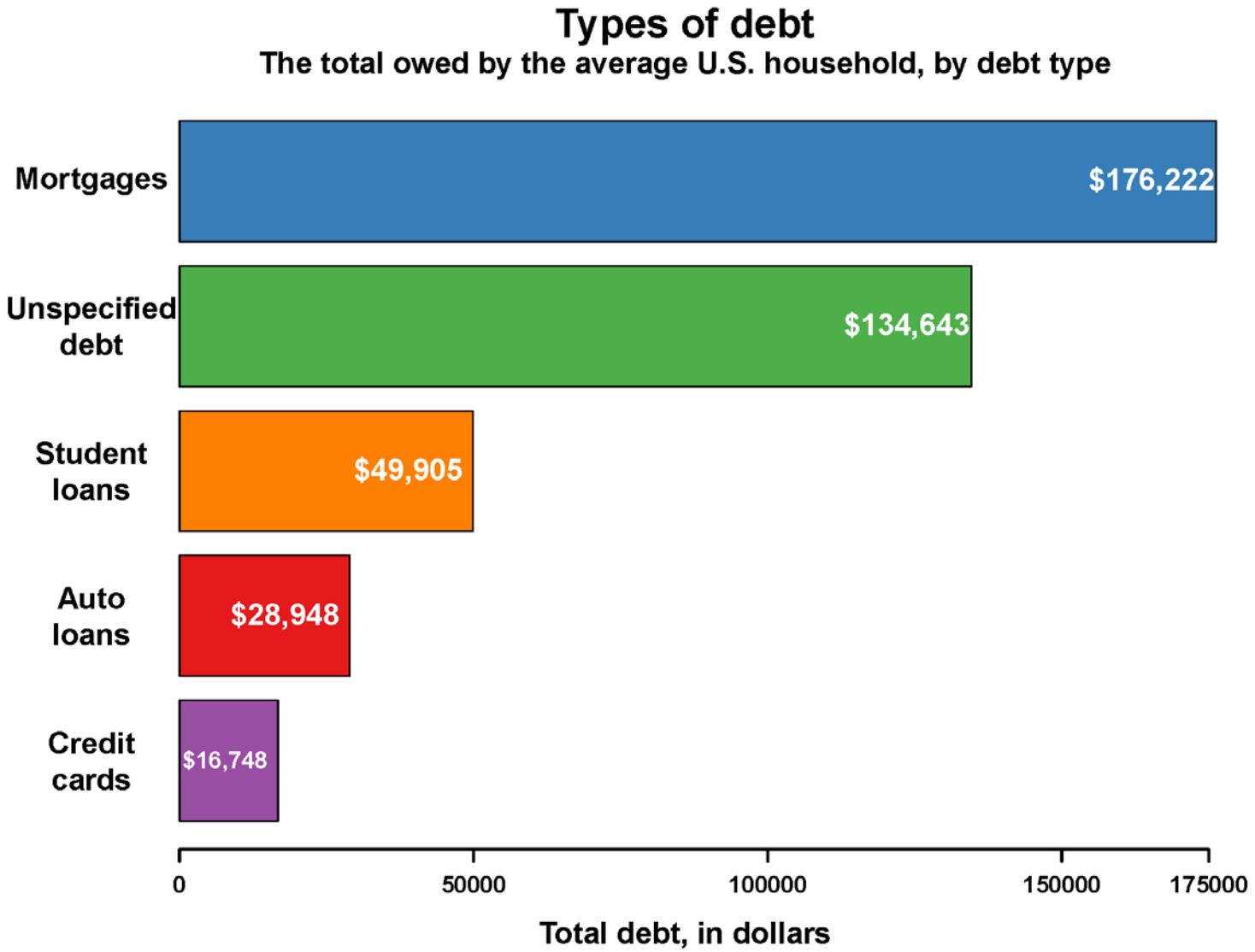


No axis labels. The x-axis has no sense of scale

Is “Any type of debt” the sum of all other types?

http://www.investmentzen.com

Example 1: A better version, created in R



Reordered bars from largest to smallest

x-axis is scaled and labeled properly

Example 2:

Ugly graph

Why is this highlighted?

You can't compare circles across columns

Legend for color of circles not easily found

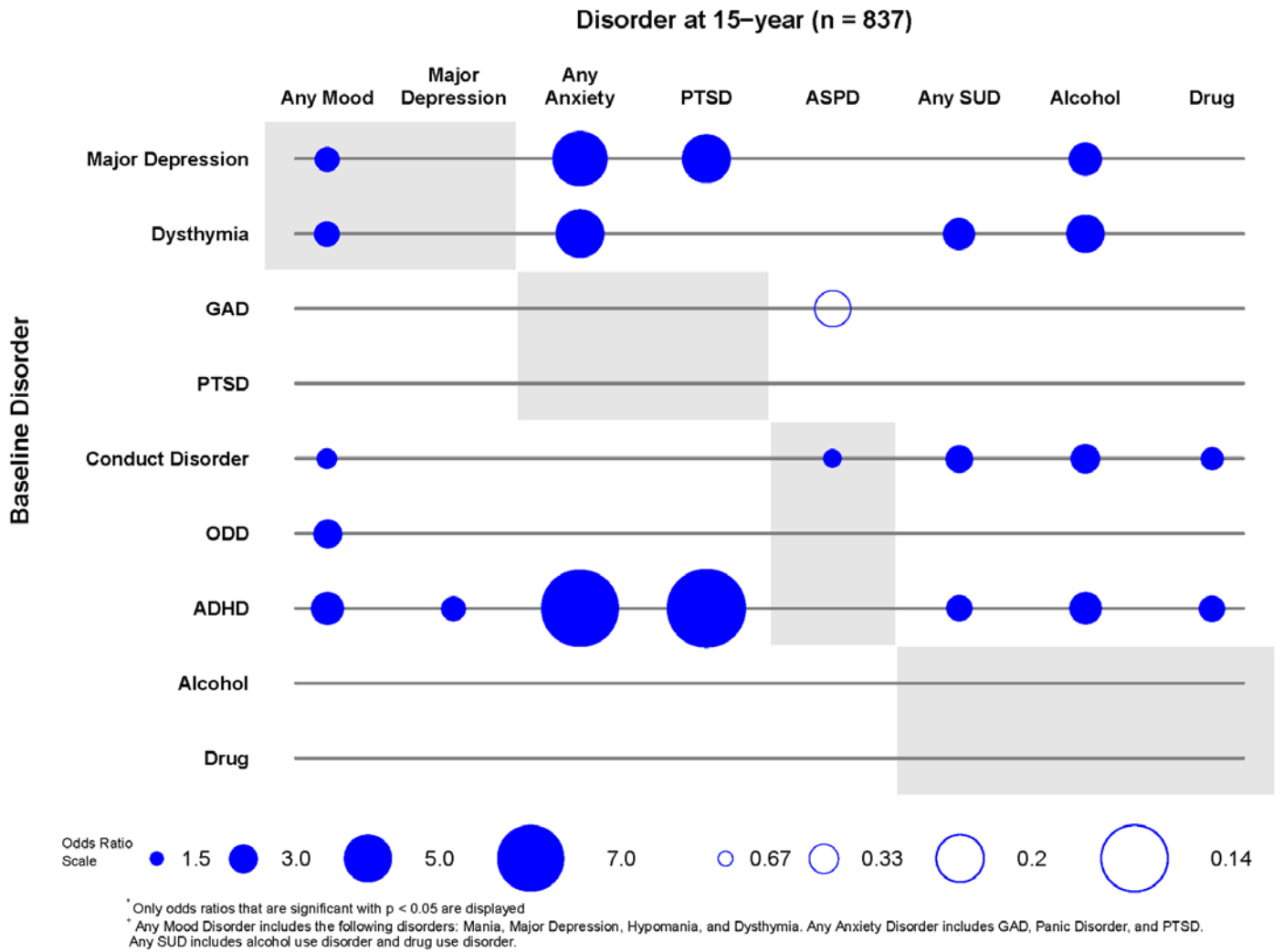
Why is this highlighted?

"Let's Talk About Sex ", New York Times, Sep 6 2008.

Example 3

Good graph

Figure 5. Odds Ratios of 15–Year DSM–IV Diagnoses Predicted From Baseline Diagnoses: Males



* Only odds ratios that are significant with $p < 0.05$ are displayed
* Any Mood Disorder includes the following disorders: Mania, Major Depression, Hypomania, and Dysthymia. Any Anxiety Disorder includes GAD, Panic Disorder, and PTSD. Any SUD includes alcohol use disorder and drug use disorder.

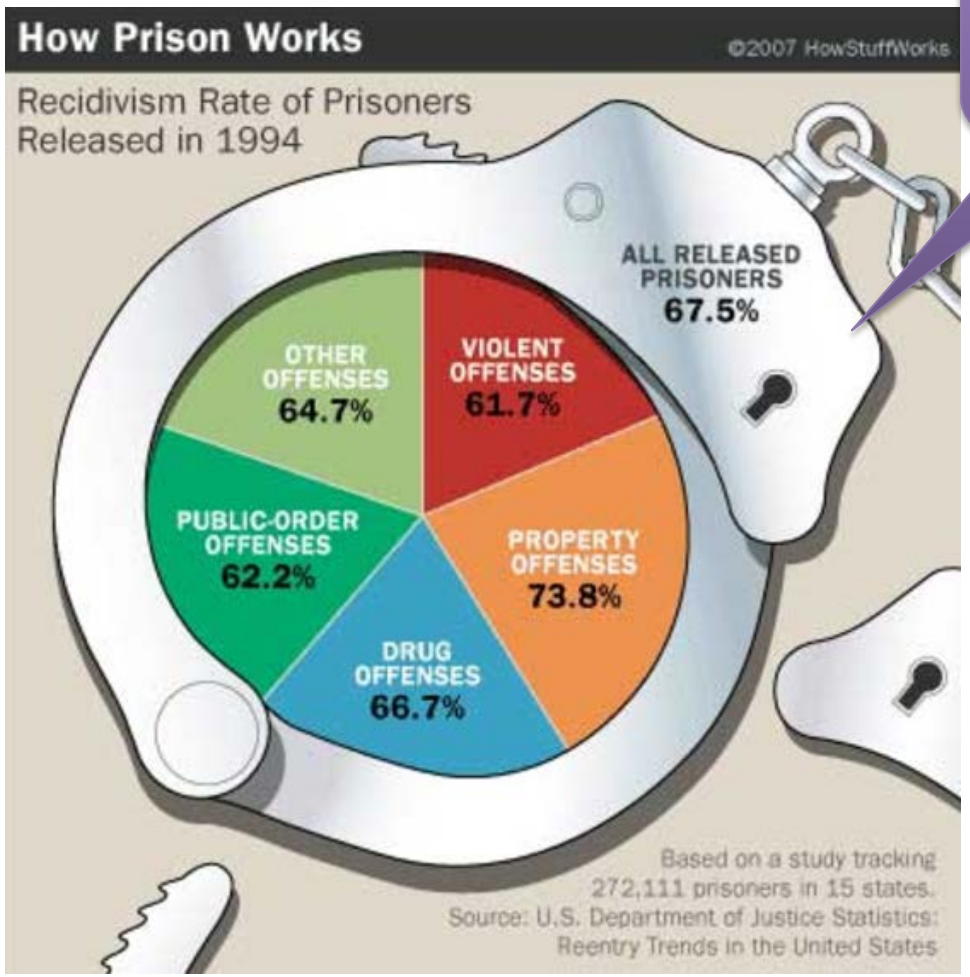
Size (and fill) of circle represents size of odds ratio

Can compare across both rows and columns

Gray boxes help group similar disorders together

Example 4

Bad and ugly graphs

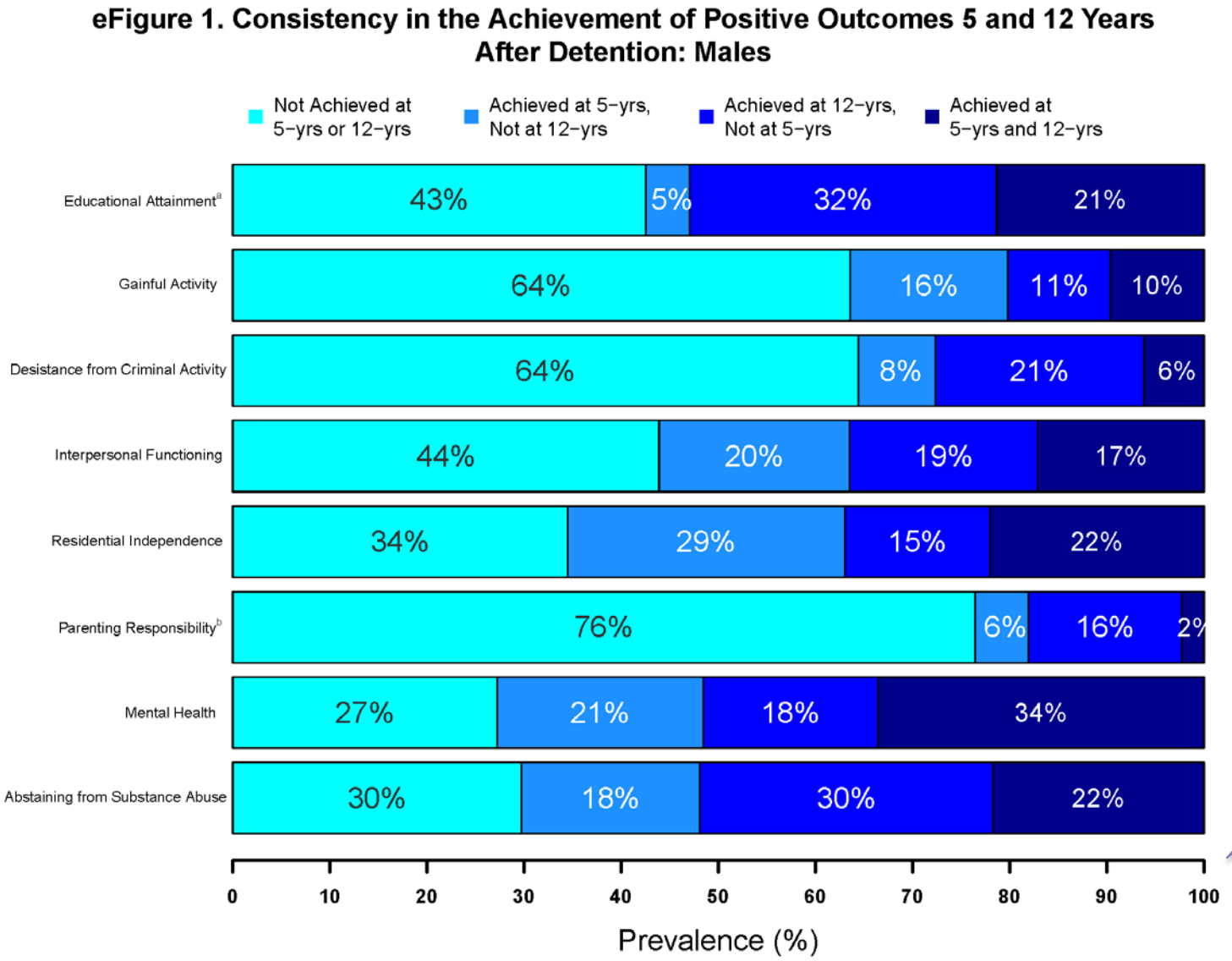


Pieces don't add up to 100%

Multiple 3D pie charts? Why?

Example 5

Good graph

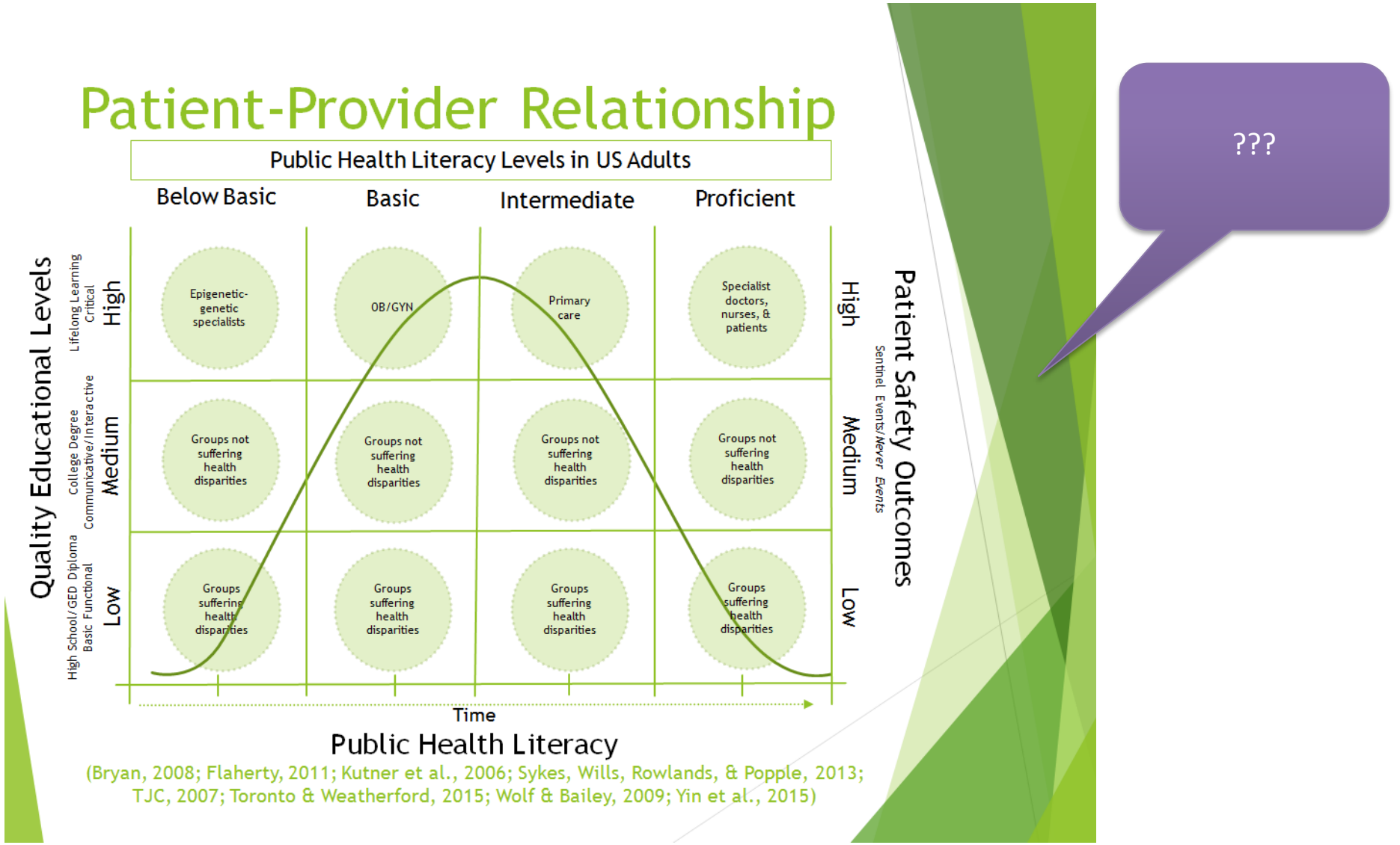


Informative colors and labels

Multiple bars are easier to read than pies

Example 6

Bad graph

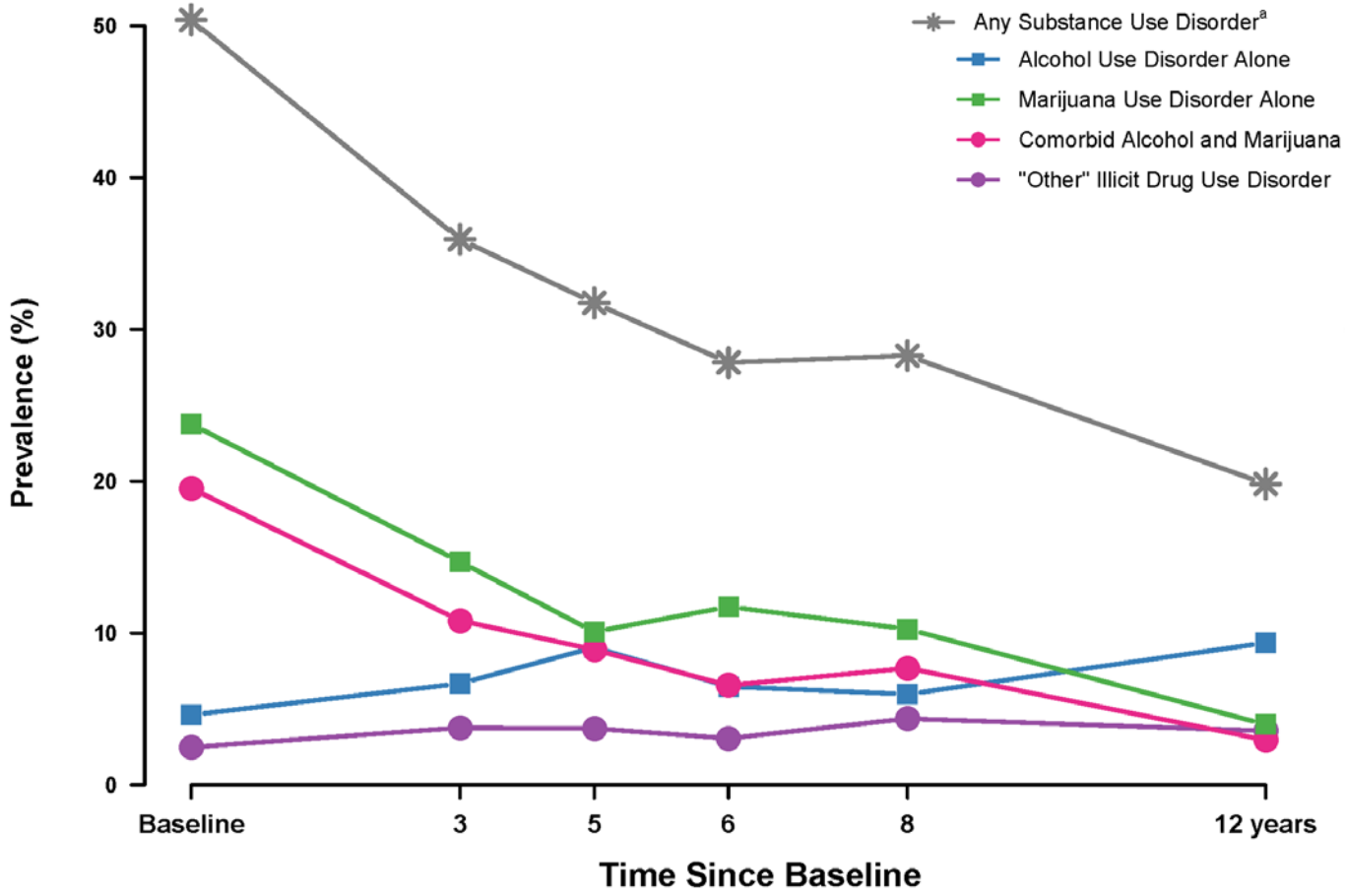


(Bryan, 2008; Flaherty, 2011; Kutner et al., 2006; Sykes, Wills, Rowlands, & Popple, 2013; TJC, 2007; Toronto & Weatherford, 2015; Wolf & Bailey, 2009; Yin et al., 2015)

Example 7

Good Graph

Figure 1. Prevalence of Substance Use Disorders During the 12 Years After Detention in Cook County (Chicago): Males (n = 1142)



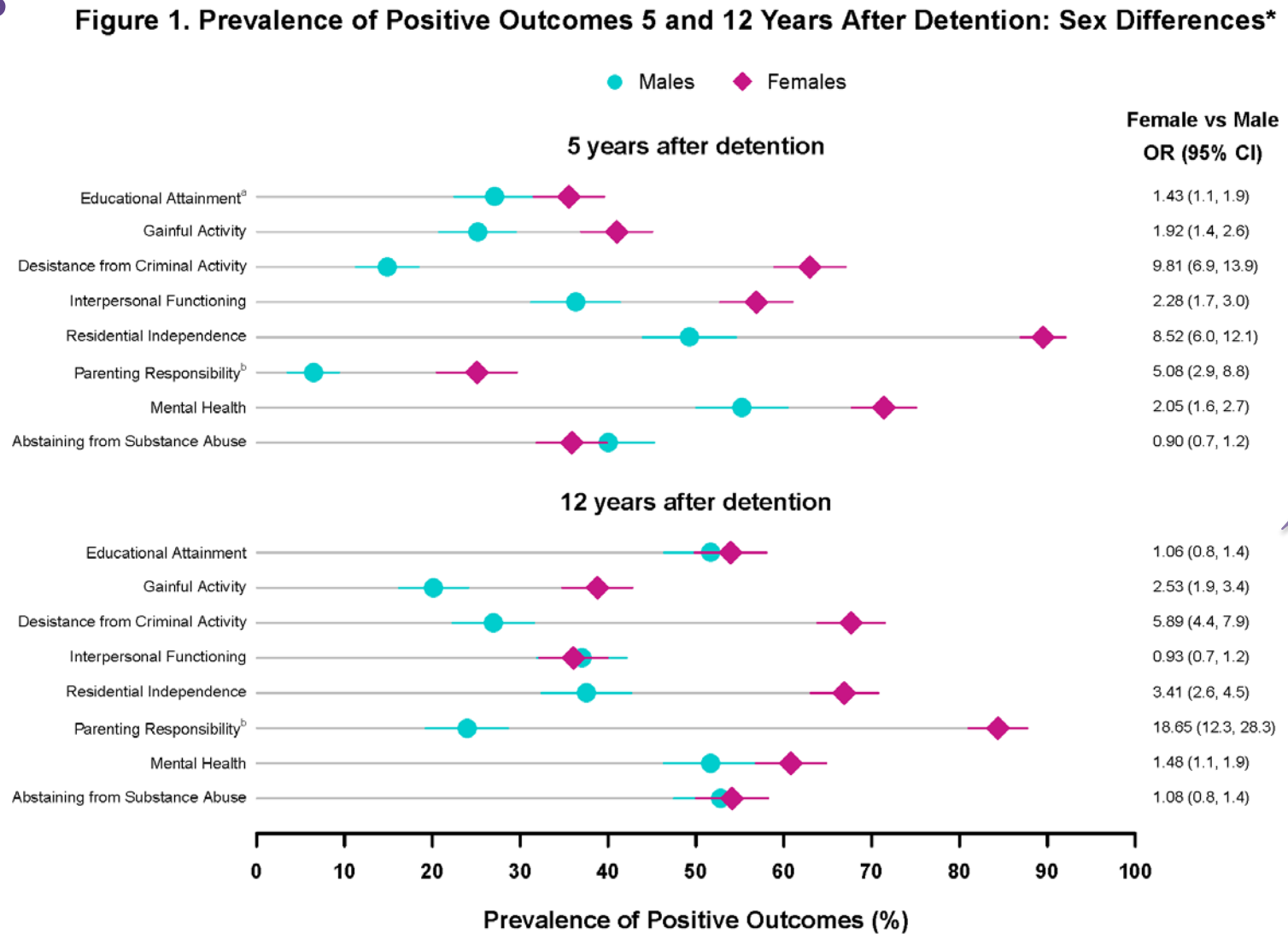
[a] Subcategories of any substance use disorder are mutually exclusive.

Simple

Good use of color and symbols

Example 8

Good graph



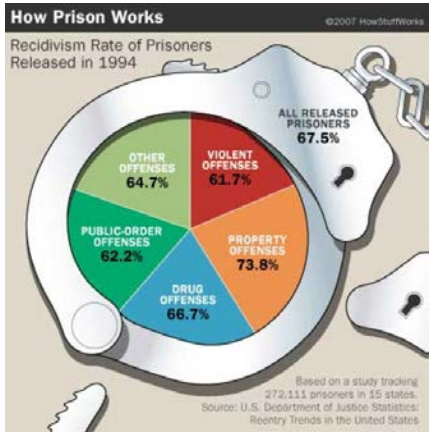
* For each positive outcome, this figure shows prevalence and associated 95% confidence intervals among males and females, and the corresponding odds ratios comparing females with males.
[a] Educational attainment excludes participants who were less than 18 years of age at the time of interview.
[b] Parenting responsibility excludes participants who did not have any children at the time of interview.

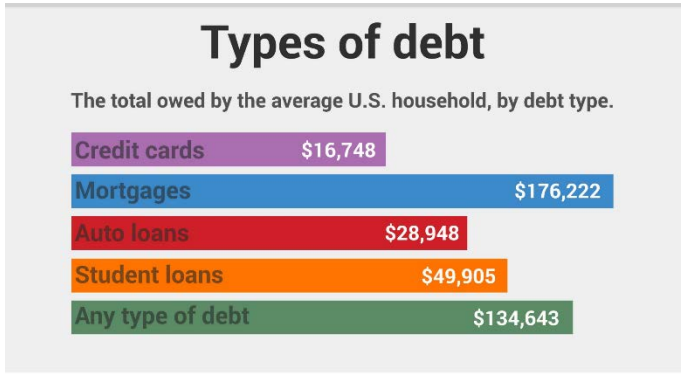
High information to ink ratio

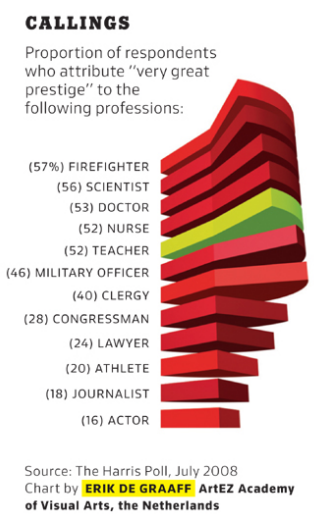
What makes a graph bad?

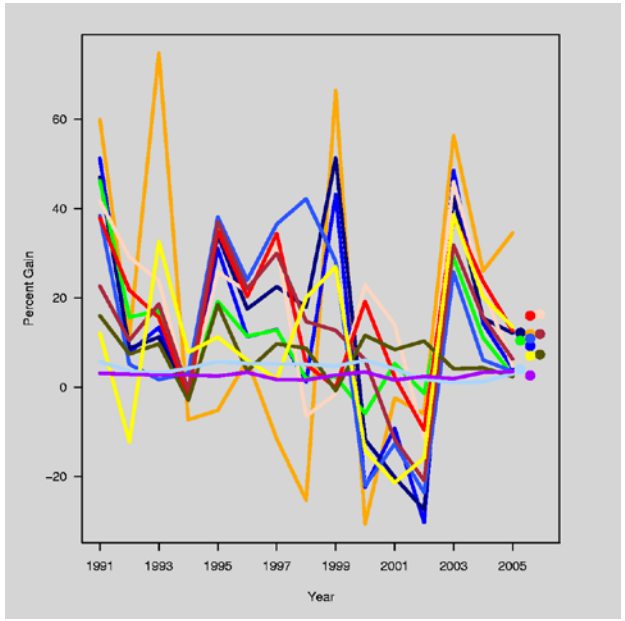
Points to keep in mind

- "Chartjunk"
  - Extraneous, distracting visual elements
- Undefined/ unlabeled axes
- Distorting the data (deliberate or accidental)
- The wrong graph for the data
- Poor choice of color





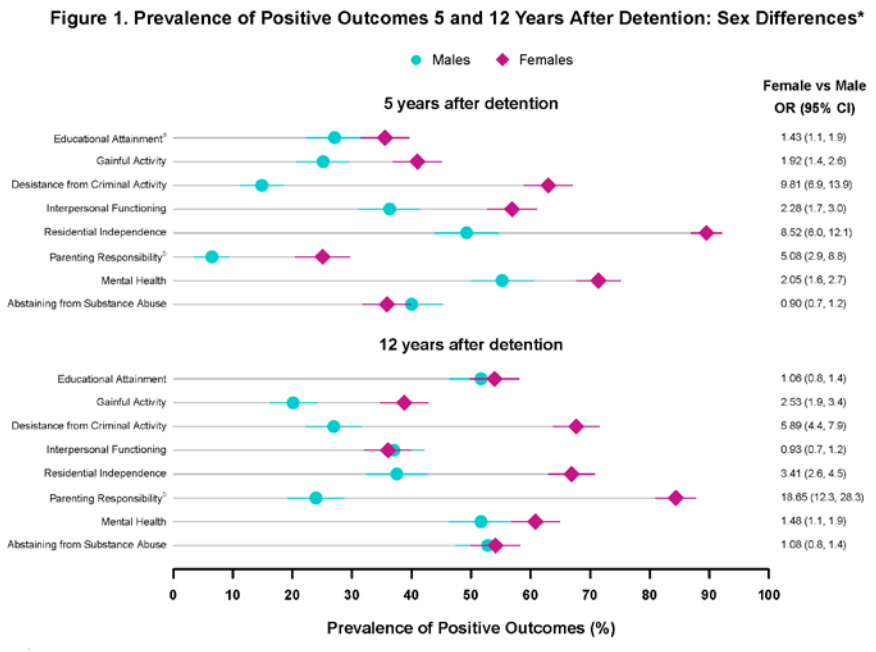


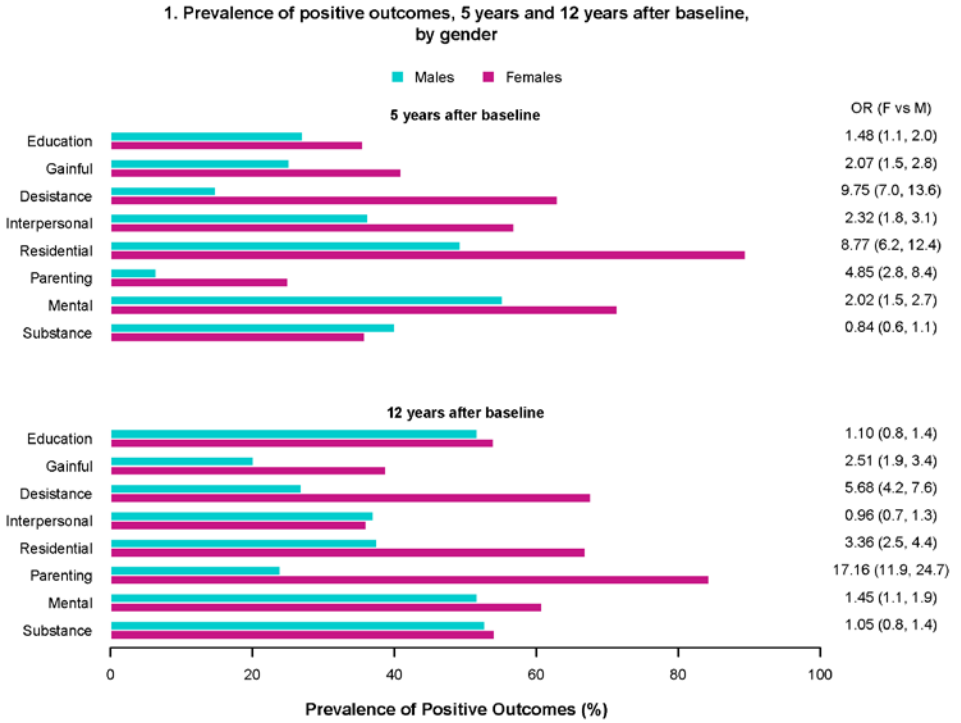


http://andrewgelman.com/2006/05/23/post_8

What makes a graph good?

Points to keep in mind: see works by Edward Tufte





- Maximum information and minimum ink
  - Present many numbers in a small space
  - Encourage the eye to compare different pieces of data
- Labels should be informative but not distracting
  - Graphics should stand on their own

- Graphics should have no more dimensions than exist in your data
  - No 3-dimensional bar plot or pie-charts
  - Only 3-d if you are plotting a surface
- Choosing the appropriate graph for your data
  - A bar plot or pie chart is not always the best choice

Northwestern Medicine®
Feinberg School of Medicine

Why should I use R?

Why can't I just use Excel for graphics?

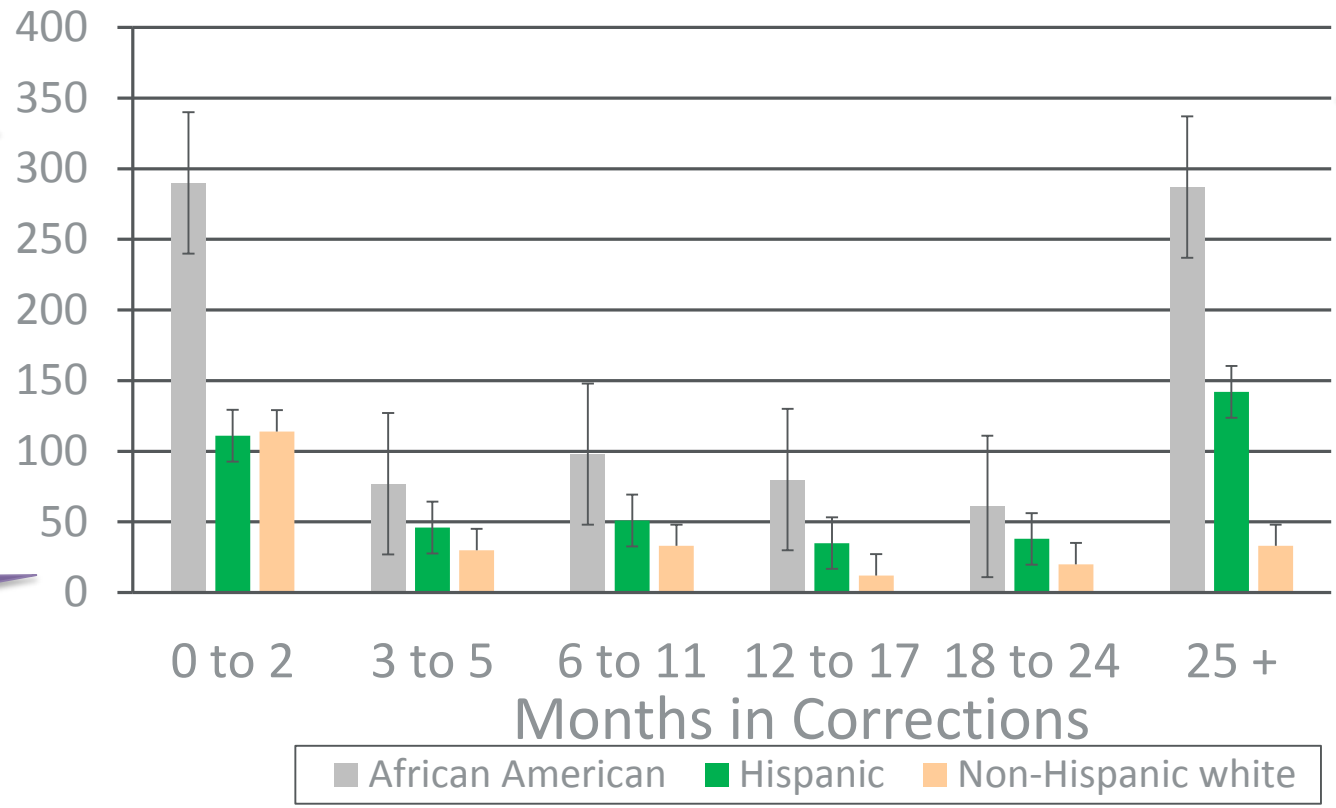
Excel can lead you astray

- Longitudinal study of juvenile delinquents (*Northwestern Juvenile Project*)
- Are there racial/ethnic differences in length of time incarcerated?

No y-axis label

"Dynamite plot"

0 much different than 1



Extra black ink is distracting

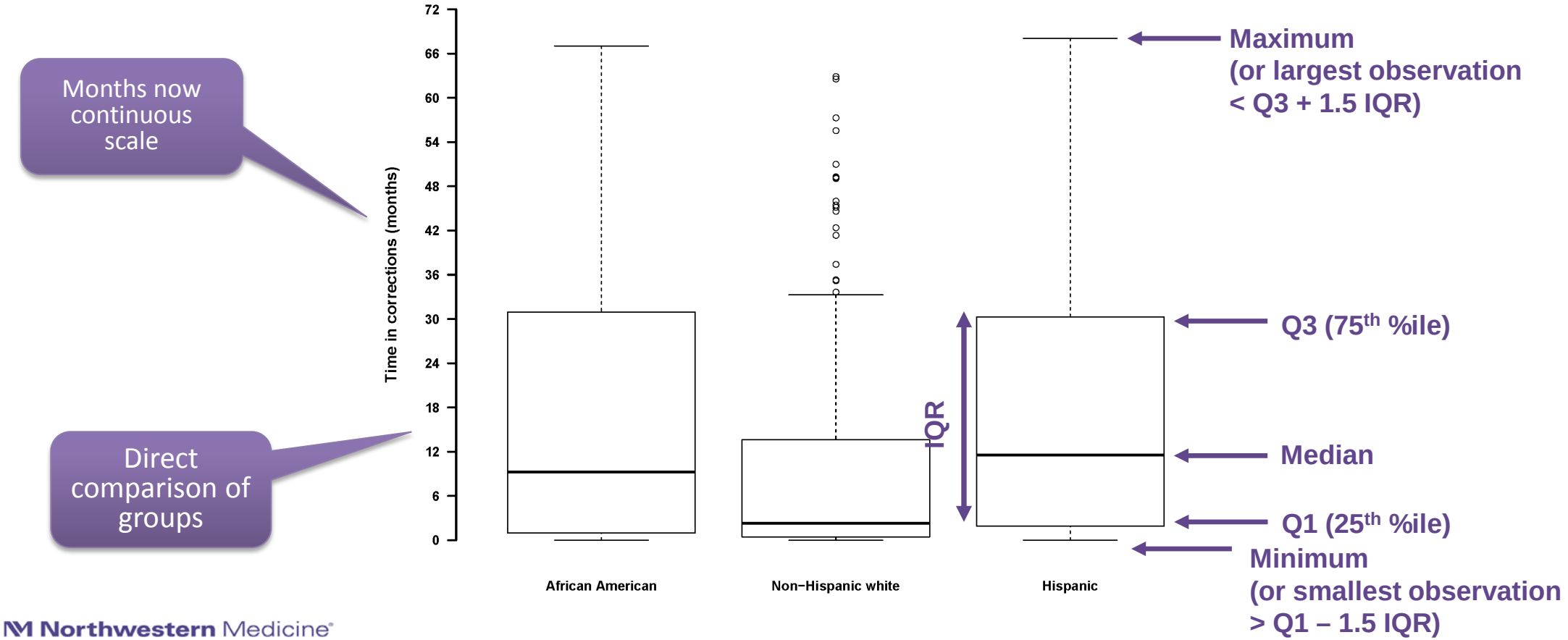
Could put actual %'s above bars

Discretizes (unequally) a continuous scale

Good Pictures: Side-by-side boxplots

Simple but conveys a lot of information

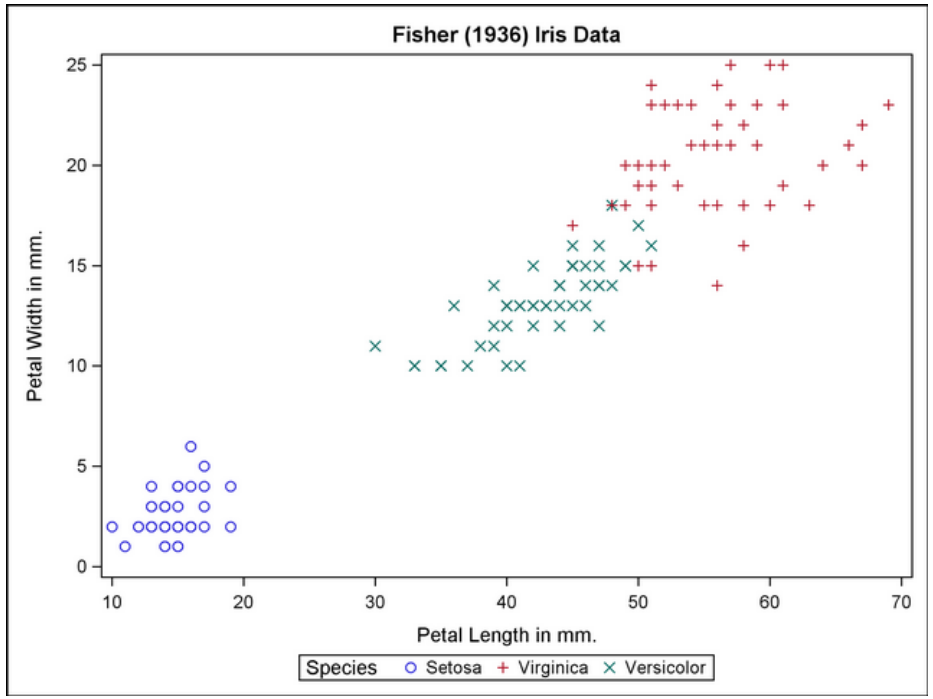
- Excel can't easily make a boxplot
- Much more appropriate picture for the data
- Conveys a lot of information in a simple way



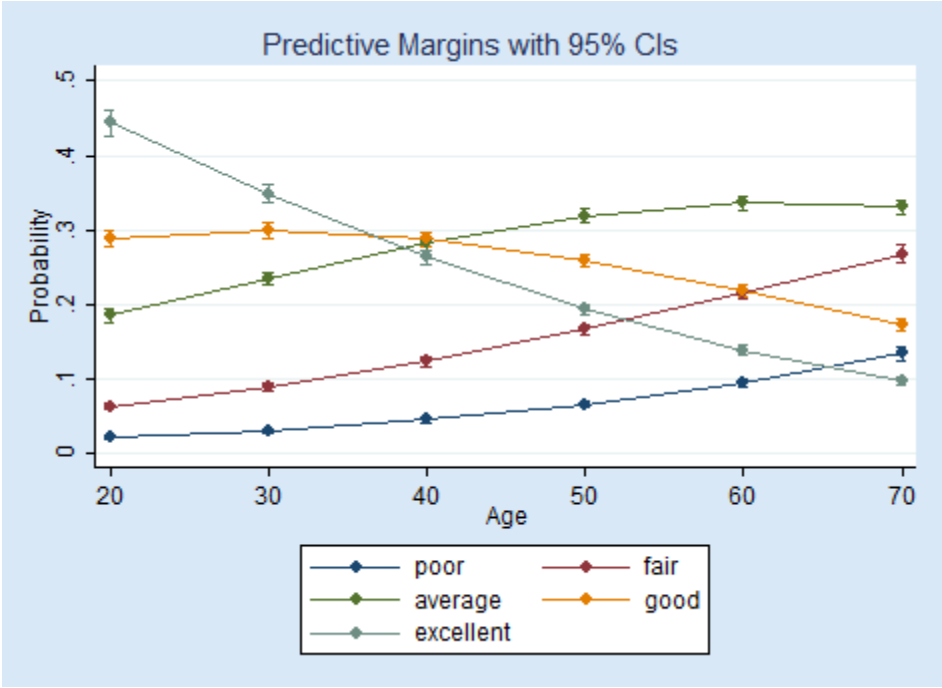
What about SAS or Stata?

Can be great, but…

- Better than Excel, can do a wider variety of plots
- Can be difficult to customize beyond default settings (and not as customizable as R)
- Expensive (R is free)



https://support.sas.com



https://www.stata.com

What is R?

- Open source programming language for statistical computing and graphics
- Provides a wide variety of statistical and graphical techniques and is highly extensible
- Similar to SAS, Stata, other statistical software
- Combine with Rstudio to give it a more user-friendly interface
- https://cran.r-project.org
- https://www.rstudio.com

R for Statistical Graphics

Why should I use R?

- Pros:
  - Reproducible, not point and click
  - Can make publication quality plots and the user is in full control of every detail
  - Can do basic plots (scatterplots, barplots, histograms) as well as custom made plots
  - Open source, free to download and use

- Cons
  - Steeper learning curve

Northwestern Medicine®
Feinberg School of Medicine

Using R for Statistical Graphics: A brief walkthrough

R for Statistical Graphics

Disclaimer!

- This is not intended to teach you all the ins and outs of R programming
  - There are plenty of resources available for that

- This is to serve as an introduction to plotting in R
  - Example code is provided

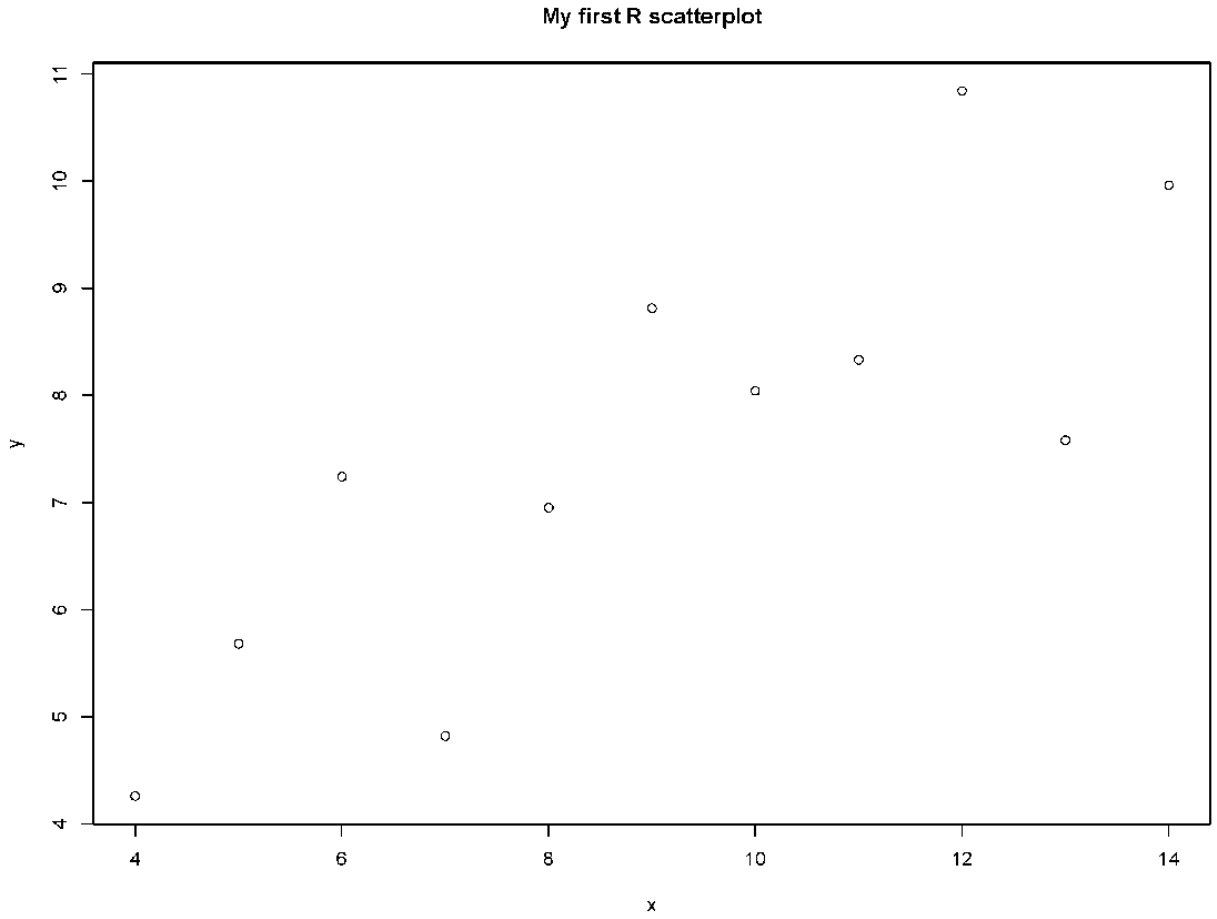
Let's make a basic plot in R

Remember Anscombe's Quartet?

```
data("anscombe")
plot(anscombe$x1, anscombe$y1,
    xlab="x", ylab="y",
    main="My first R scatterplot")
```

Bare bones, default settings

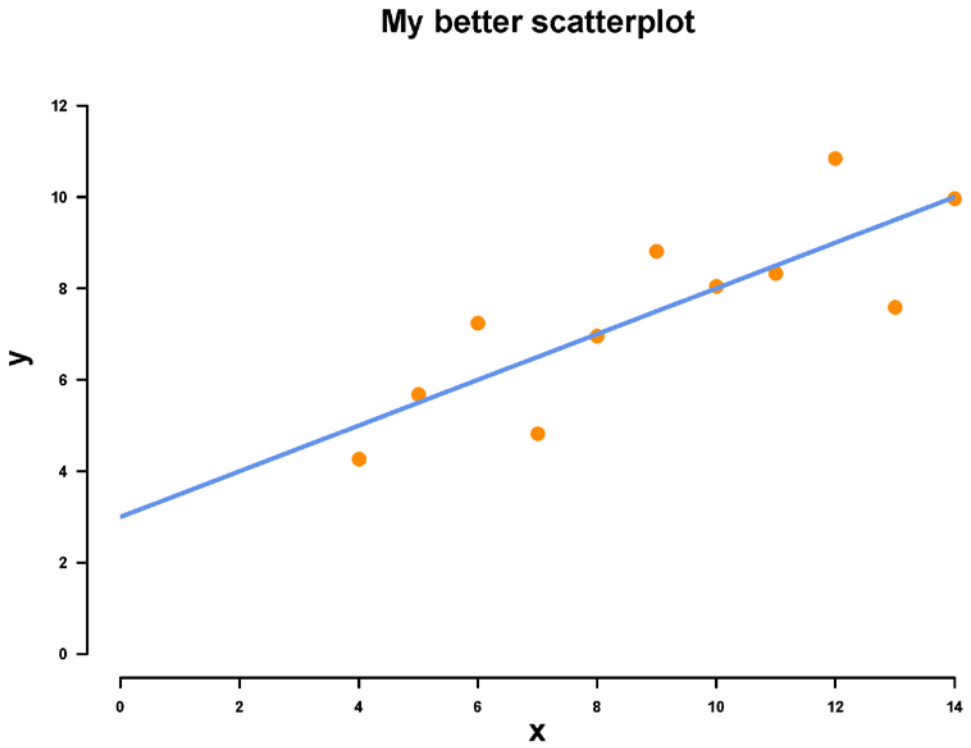
How can we make it better?



Let's make a plot in R

Let's make it look nicer!

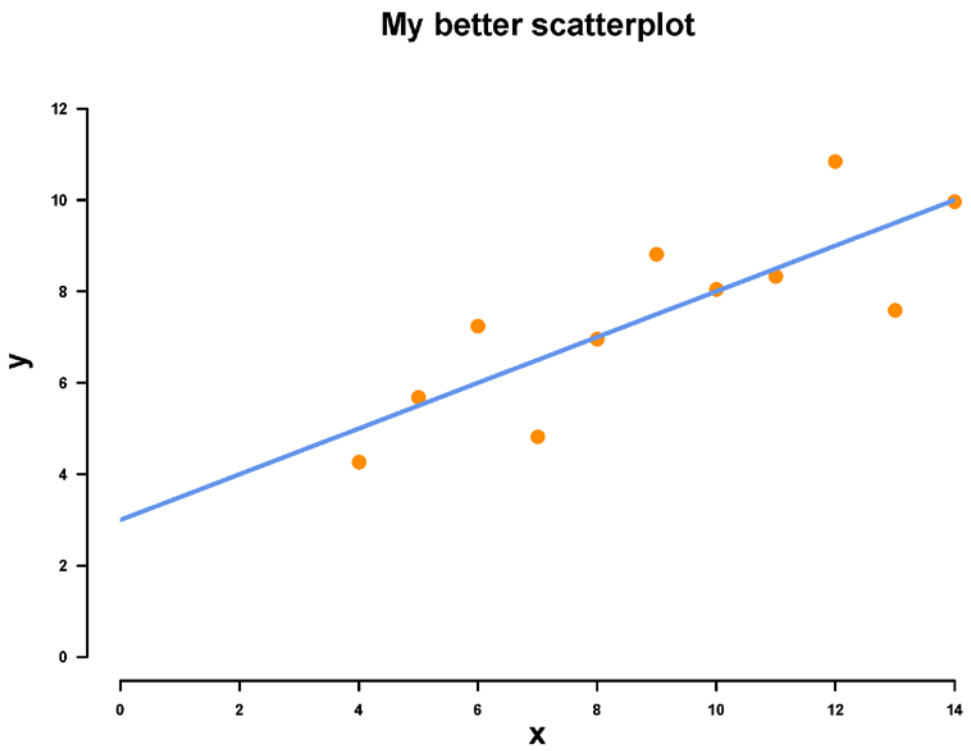
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

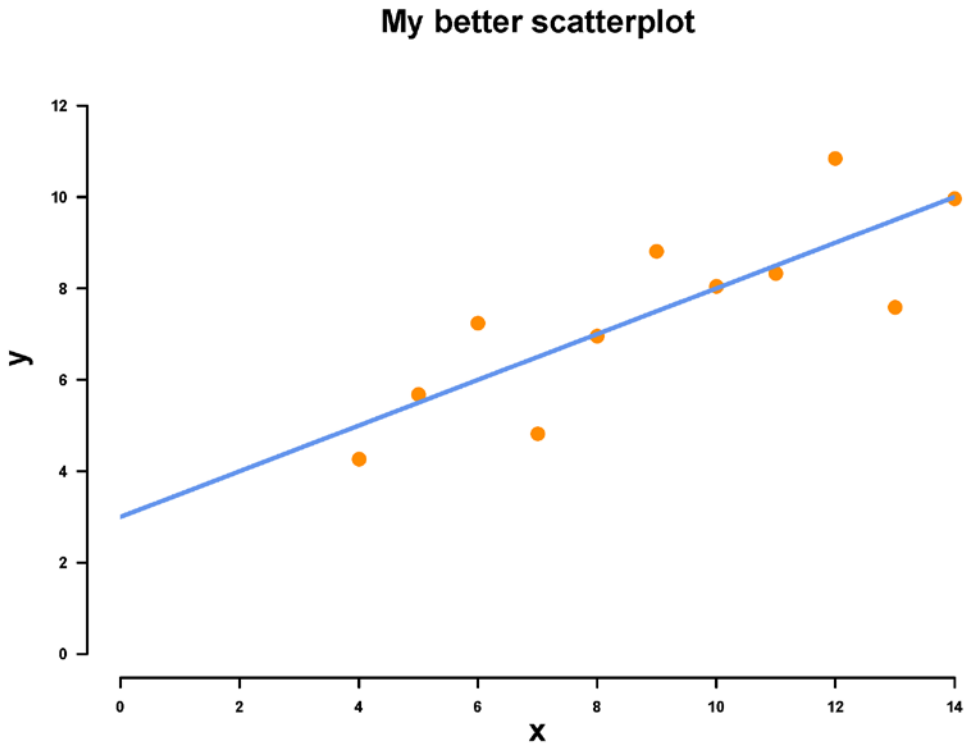
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



File types for Statistical Graphics

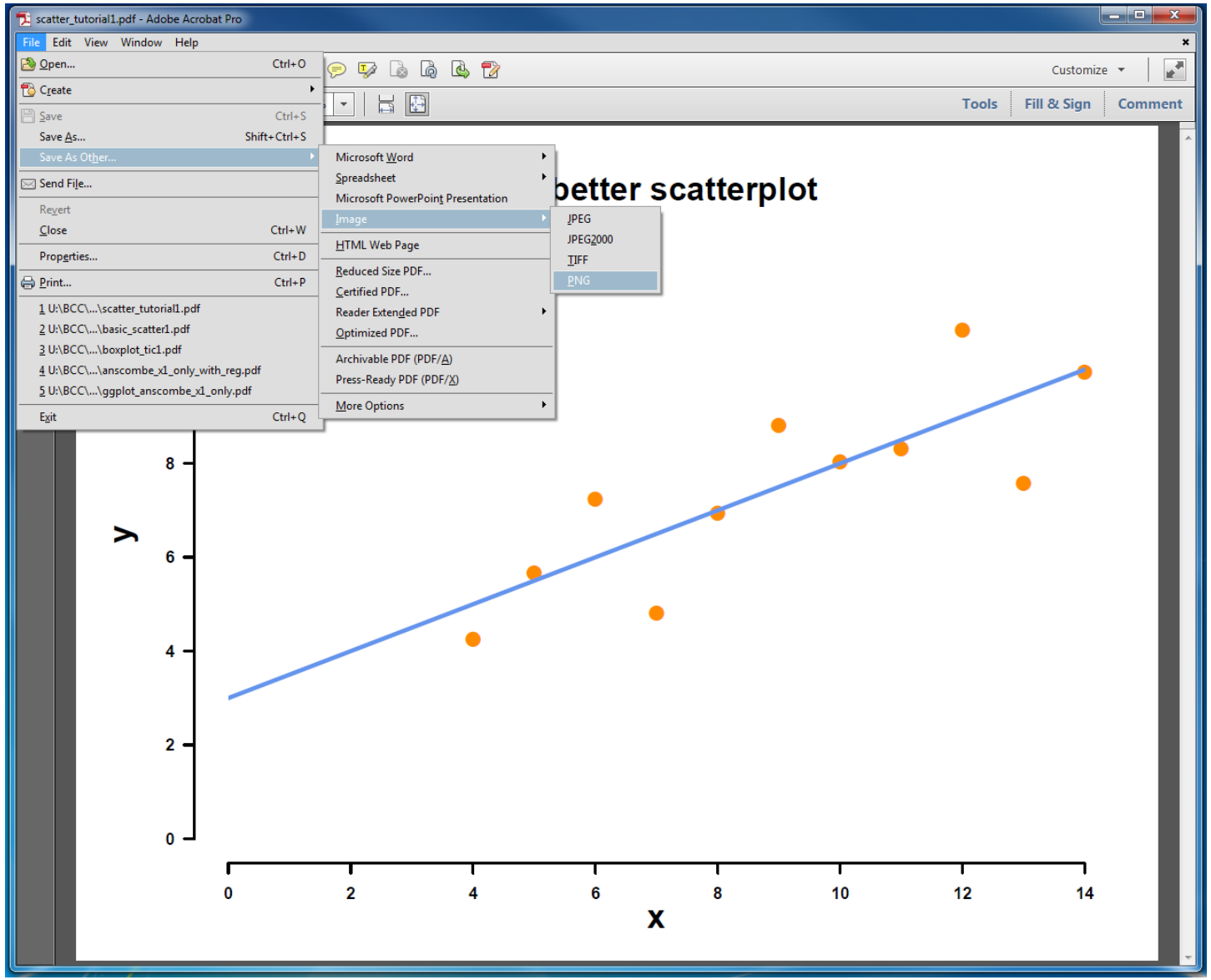
Which is best?

- Vector based graphics
  - Constructed using mathematical formulas
  - High quality graphics
  - Zoom in without any degradation in image quality
  - Best: save graphics as .pdf
- Raster images
  - Used colored pixels to form an image
  - Cannot be resized without compromising resolution
  - Can look grainy and distorted
  - Best: save graphics as .png
  - Worst: .jpg or .gif (great for photos, not for statistical graphics)

PDF to PNG

How to

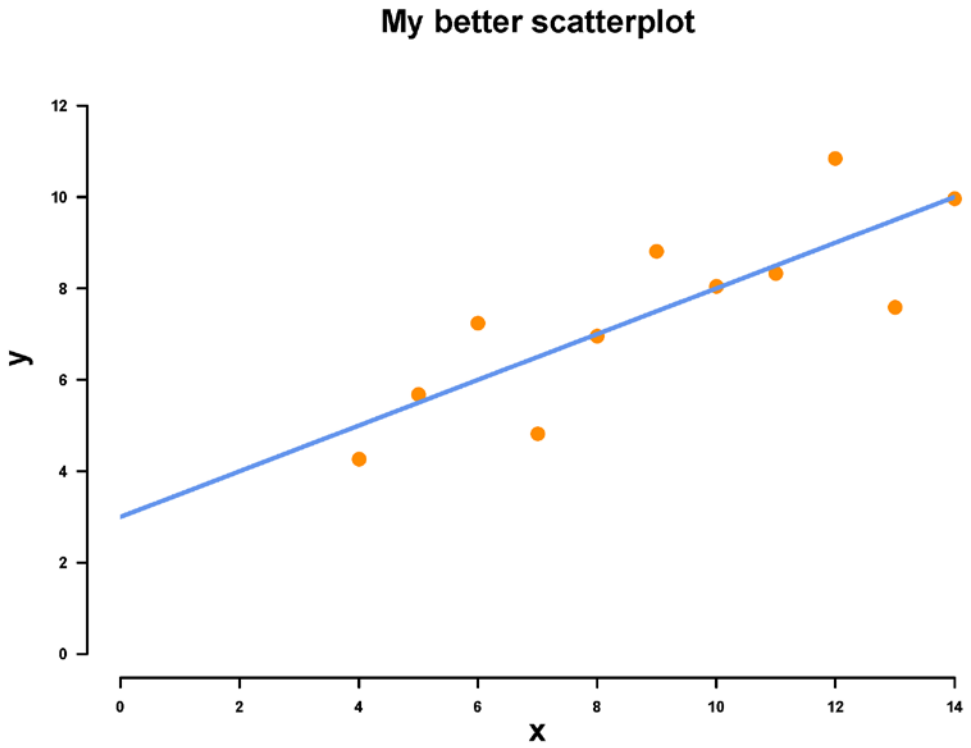
- Can be hard to insert pdf into Word doc or Powerpoint
- One solution: Save as .png file



Let's make a plot in R

Let's make it look nicer!

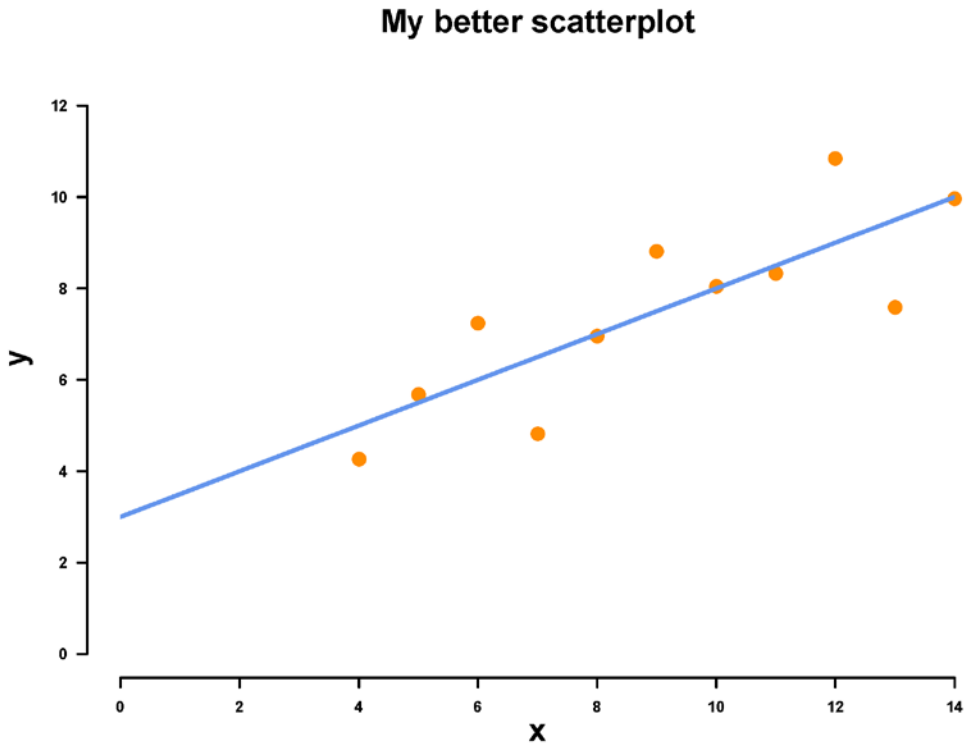
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
     width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

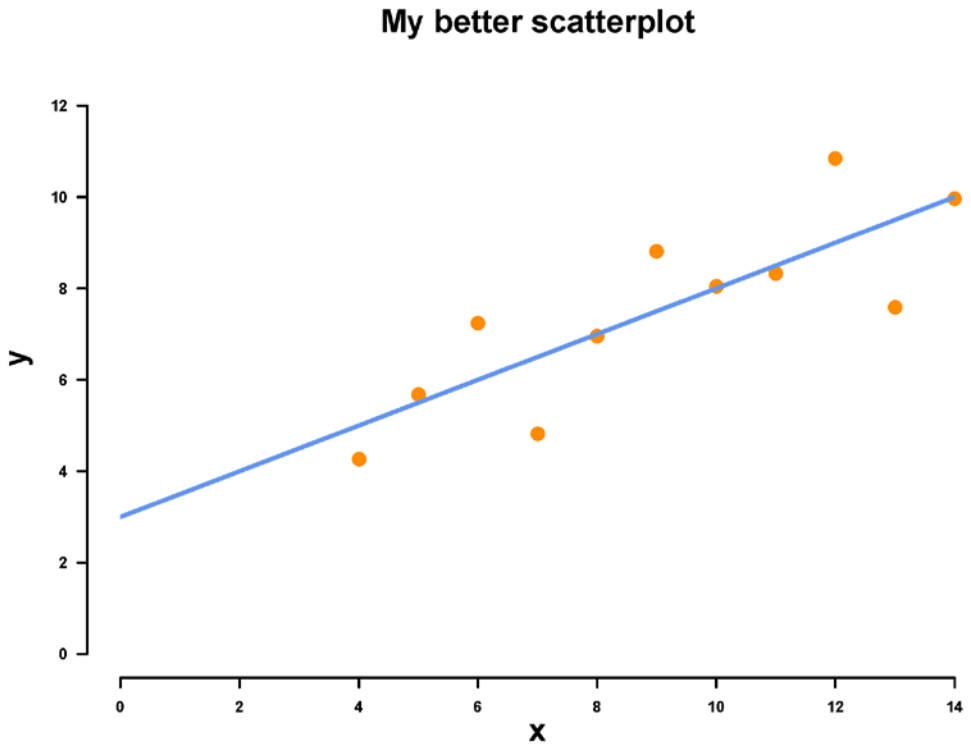
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

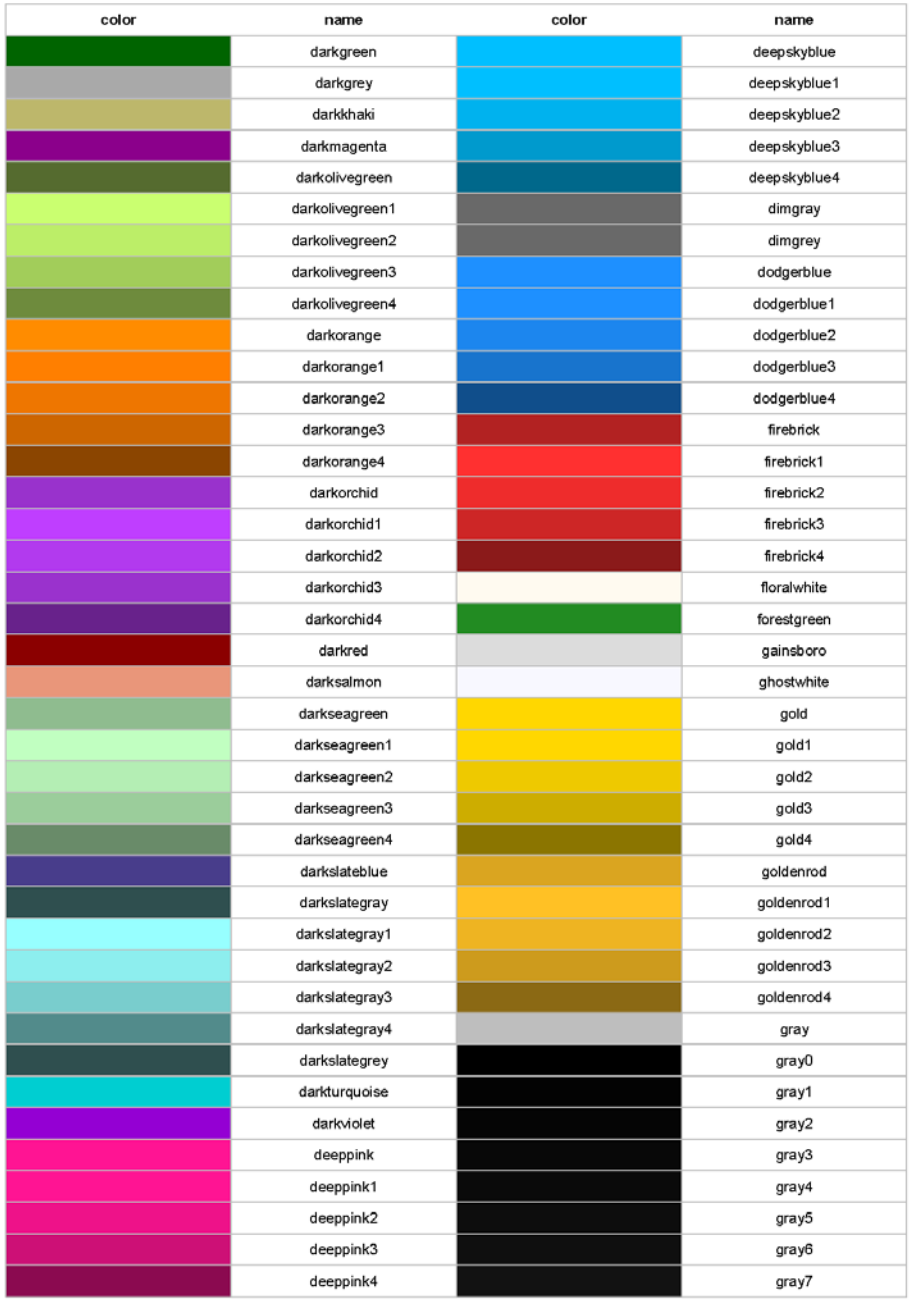
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Statistical Graphics: Colors

How to choose appropriate colors in R

- Named colors in R
  - Hundreds to choose from
  - Color should not be gratuitous
  - http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf

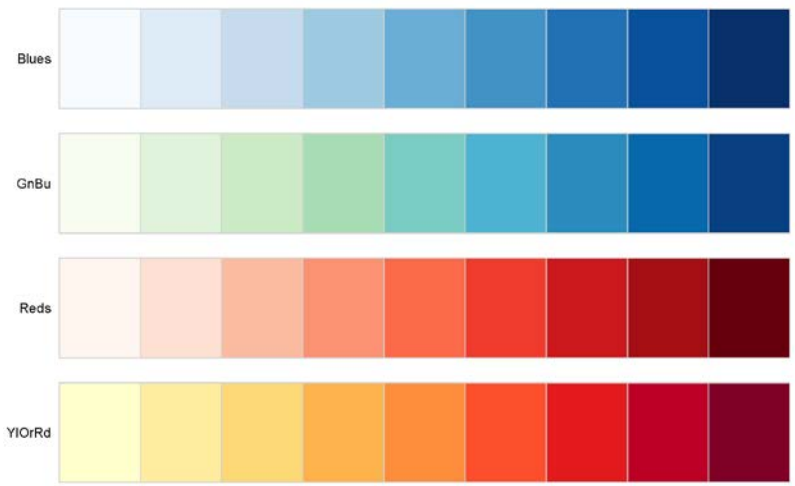


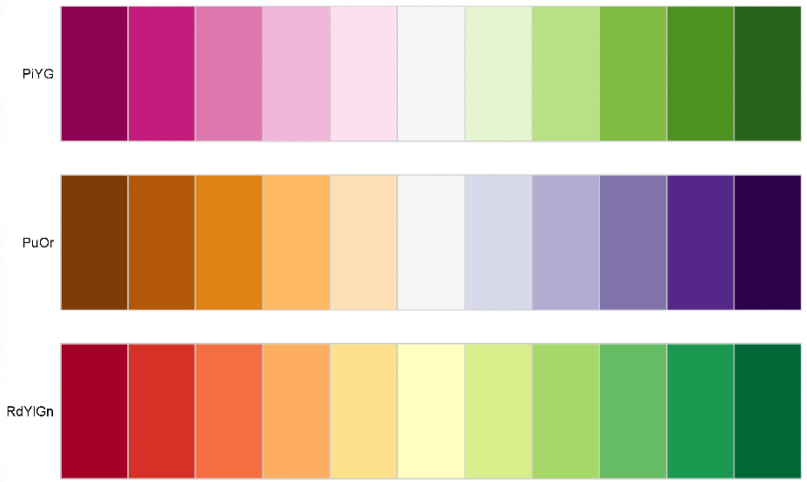
color	name	color	name
	darkgreen		deepskyblue
	darkgrey		deepskyblue1
	darkkhaki		deepskyblue2
	darkmagenta		deepskyblue3
	darkolivegreen		deepskyblue4
	darkolivegreen1		dimgray
	darkolivegreen2		dimgrey
	darkolivegreen3		dodgerblue
	darkolivegreen4		dodgerblue1
	darkorange		dodgerblue2
	darkorange1		dodgerblue3
	darkorange2		dodgerblue4
	darkorange3		firebrick
	darkorange4		firebrick1
	darkorchid		firebrick2
	darkorchid1		firebrick3
	darkorchid2		firebrick4
	darkorchid3		floralwhite
	darkorchid4		forestgreen
	darkred		gainsboro
	darksalmon		ghostwhite
	darkseagreen		gold
	darkseagreen1		gold1
	darkseagreen2		gold2
	darkseagreen3		gold3
	darkseagreen4		gold4
	darkslateblue		goldenrod
	darkslategray		goldenrod1
	darkslategray1		goldenrod2
	darkslategray2		goldenrod3
	darkslategray3		goldenrod4
	darkslategray4		gray
	darkslategrey		gray0
	darkturquoise		gray1
	darkviolet		gray2
	deeppink		gray3
	deeppink1		gray4
	deeppink2		gray5
	deeppink3		gray6
	deeppink4		gray7

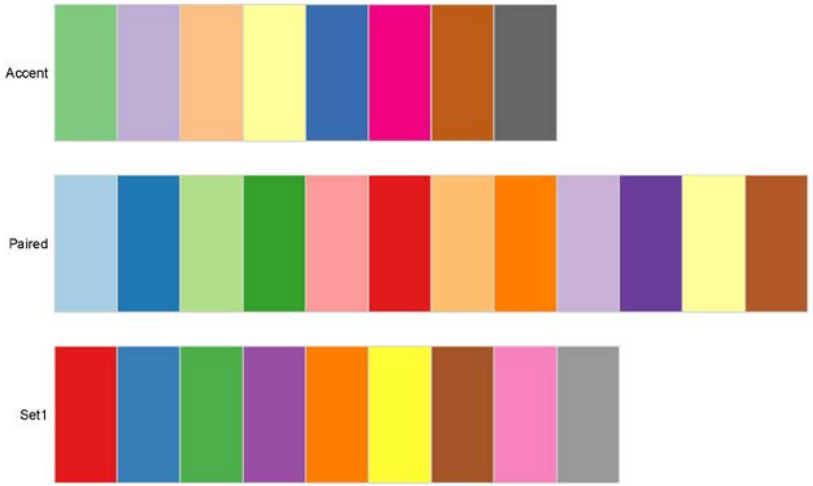
Statistical Graphics: Colors

RcolorBrewer: R package with preset color palettes

- **Sequential palettes**: suited to ordered data that progress from low to high
- **Diverging palettes**: equal emphasis on mid-range critical values and extremes at both ends of data range
- **Qualitative palettes**: used to create primary visual differences between classes



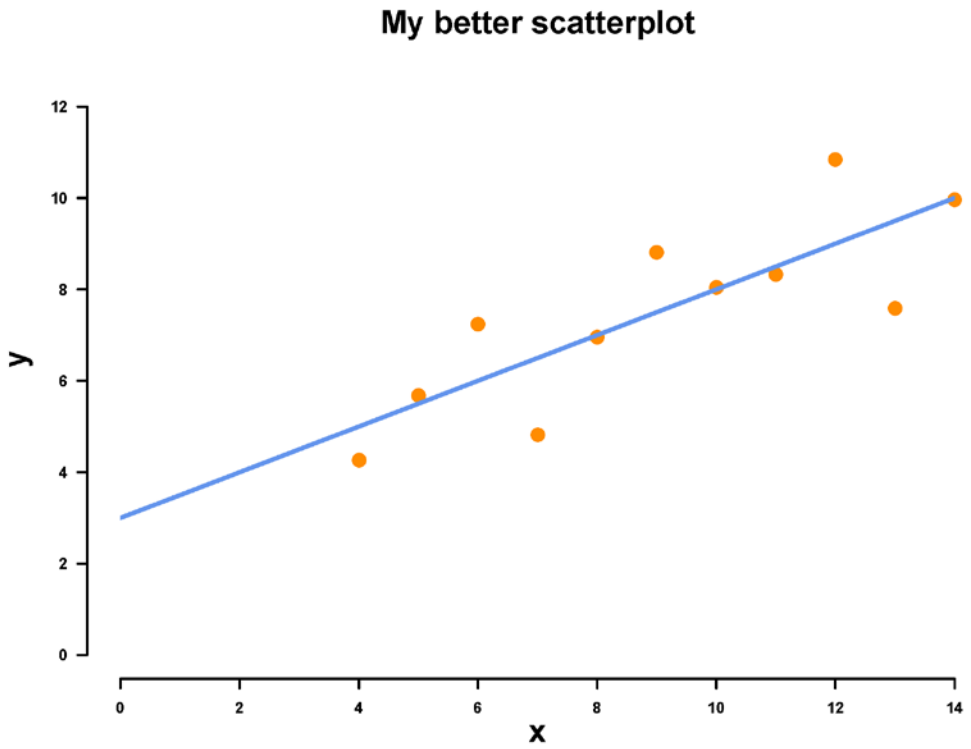




Let's make a plot in R

Let's make it look nicer!

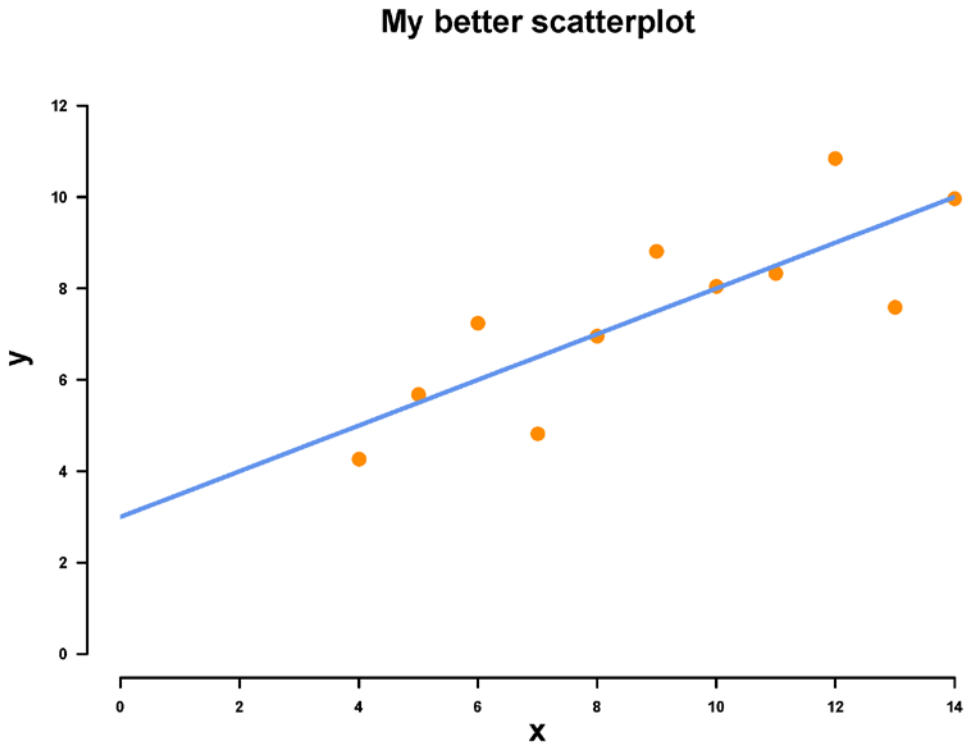
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

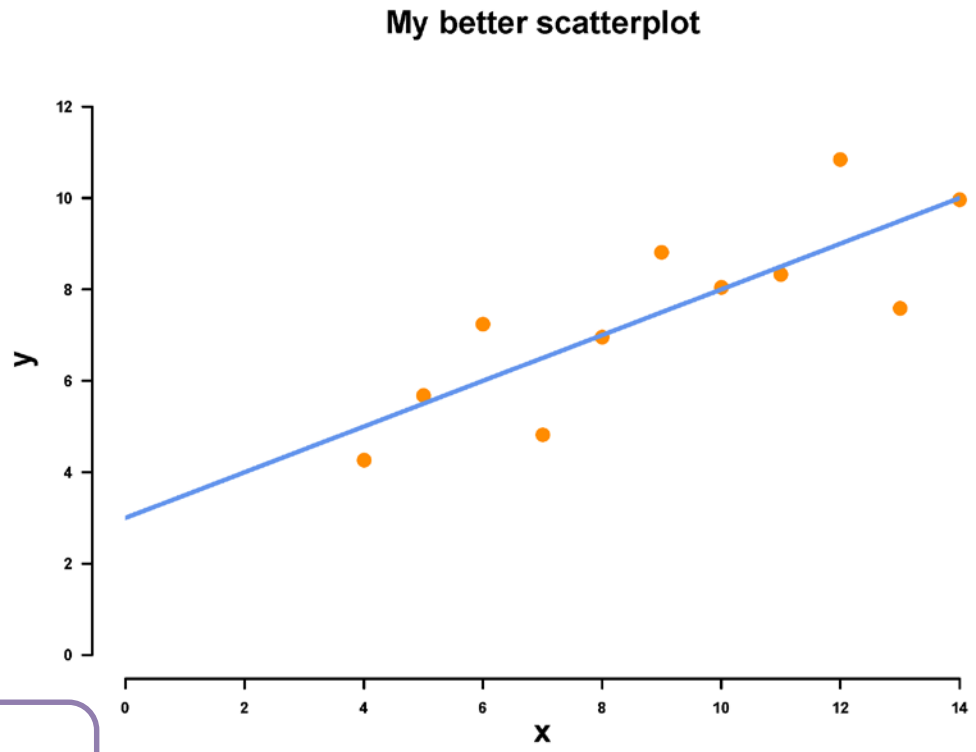
```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
     width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```



Let's make a plot in R

Let's make it look nicer!

```
data("anscombe")
pdf("C:/Users/daa745/Documents/R/scatter1.pdf",
    width=11,height=8.5)
par(mar=c(5,6,4,2))
plot(anscombe$x1, anscombe$y1,
     xlim=c(0,14), ylim=c(0,13),
     xlab="", ylab="",
     xaxt="n", yaxt="n", bty="n",
     pch=16, col="darkorange", cex=2)
clip(0,14,0,12)
abline(a=3, b=.5, col="cornflowerblue", lwd=4)
axis(1, at=seq(0,14,2), cex=1, cex.axis=1, lwd=2, font=2)
axis(2, at=seq(0,12,2), cex=1, cex.axis=1, lwd=2, font=2, las=1)
mtext("x", side=1, outer=F, cex=2, line=2.5, font=2)
mtext("y", side=2, outer=F, cex=2, line=3, font=2)
mtext("My better scatterplot", side=3, outer=F, cex=2, line=0, font=2)
dev.off()
```

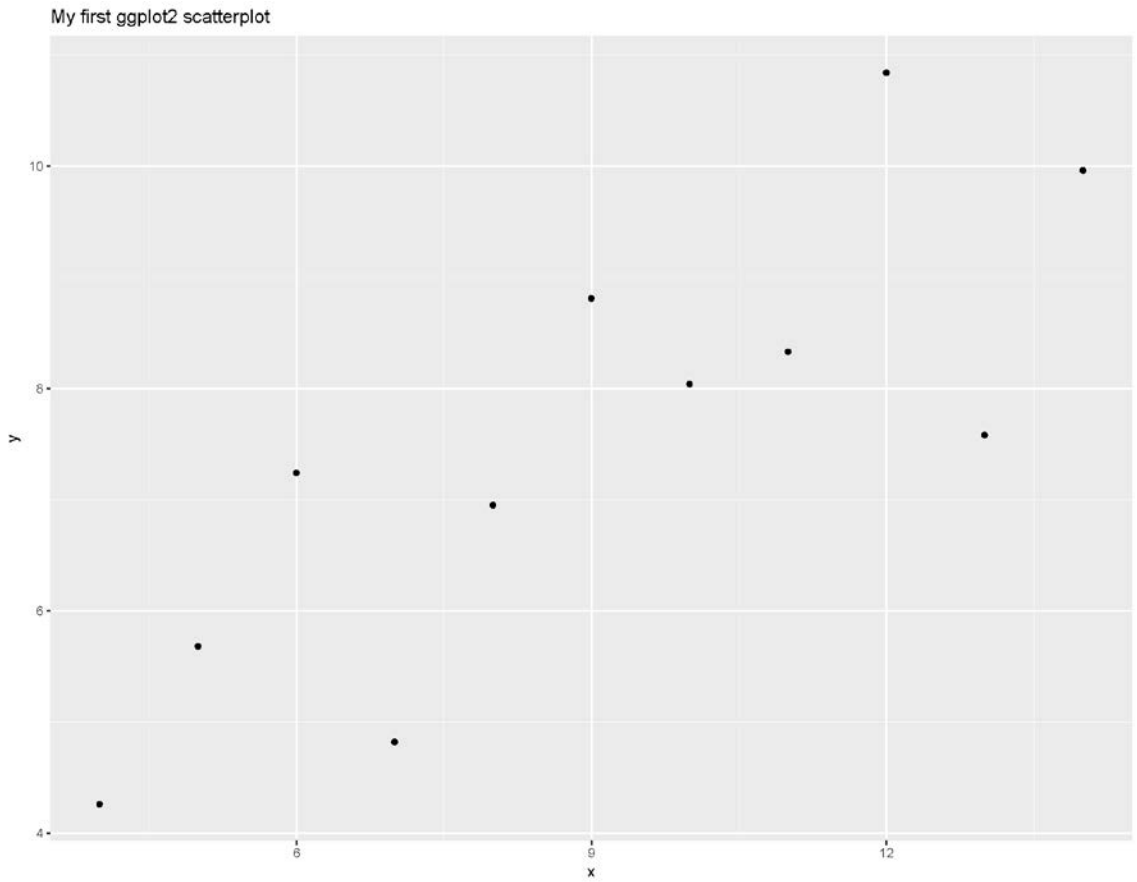


Alternative R plotting functions: ggplot2

But wait, there's more!

- Alternative, powerful graphical plotting system in R
- Requires you to install ggplot2 library

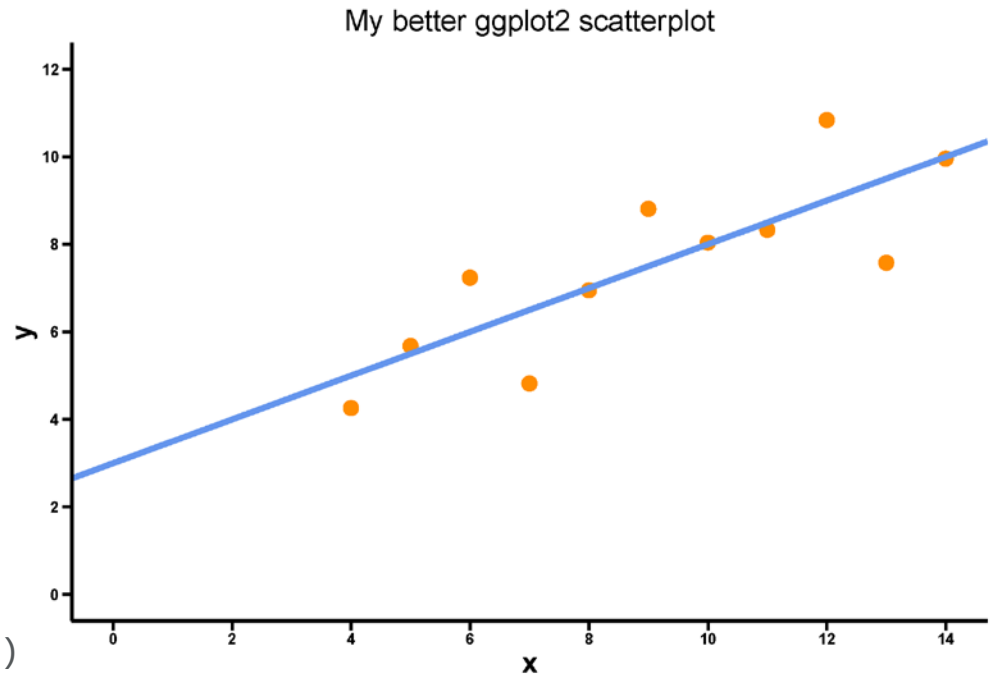
```
p1 <- ggplot(anscombe) + geom_point(aes(x1, y1)) +
   labs(title = "My first ggplot2 scatterplot",
   x = "x",
   y ="y")
p1
```



Let's make a ggplot in R

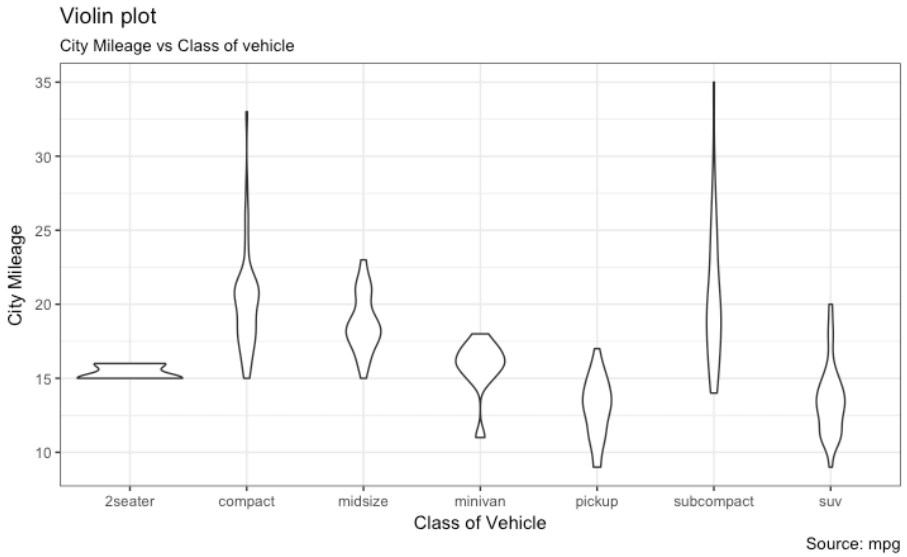
Let's make it look like our previous "nice" plot

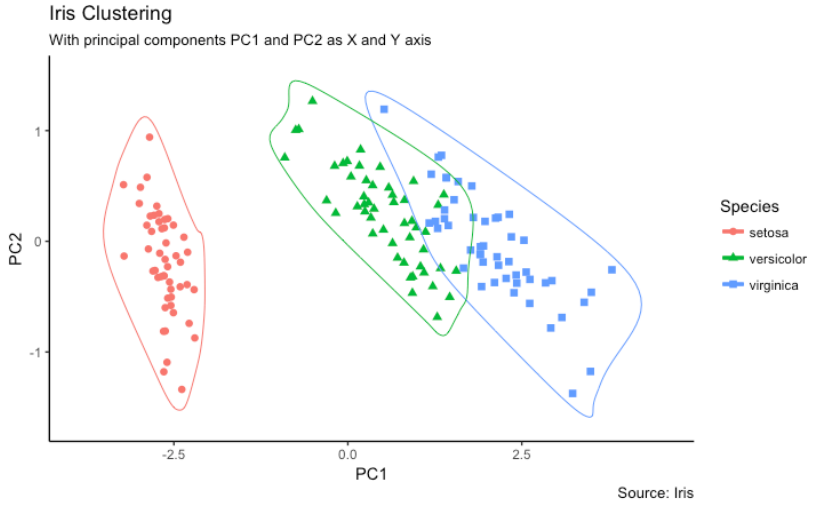
```
p1 <- ggplot(anscombe) + geom_point(aes(x1, y1), color = "darkorange", size = 5) +
 theme(
  axis.line = element_line(color="black", size=1.2),
  axis.text = element_text(color="black", face="bold", size=12),
  axis.ticks = element_line(color="black", size=1),
  axis.ticks.length = unit(.25, "cm"),
  panel.background = element_rect(fill="white", color="white"),
  axis.title = element_text(color="black", face="bold", size=22),
  plot.margin = unit(c(3,1,1,1), "cm"),
  plot.title = element_text(hjust=0.5, size=22)
  ) +
 geom_abline(intercept = 3, slope = 0.5,
   color = "cornflowerblue", size=2) +
 scale_x_continuous(name = "x", limits = c(0, 14),
   breaks=seq(0,14,2)) +
 scale_y_continuous(name = "y", limits = c(0,12),
   breaks=seq(0,12,2)) +
 ggtitle("My better ggplot2 scatterplot")

pdf("C:/Users/daa745/Documents/R/gg_scatter1.pdf",width=11,height=8.5)
print(p1)
dev.off()
```

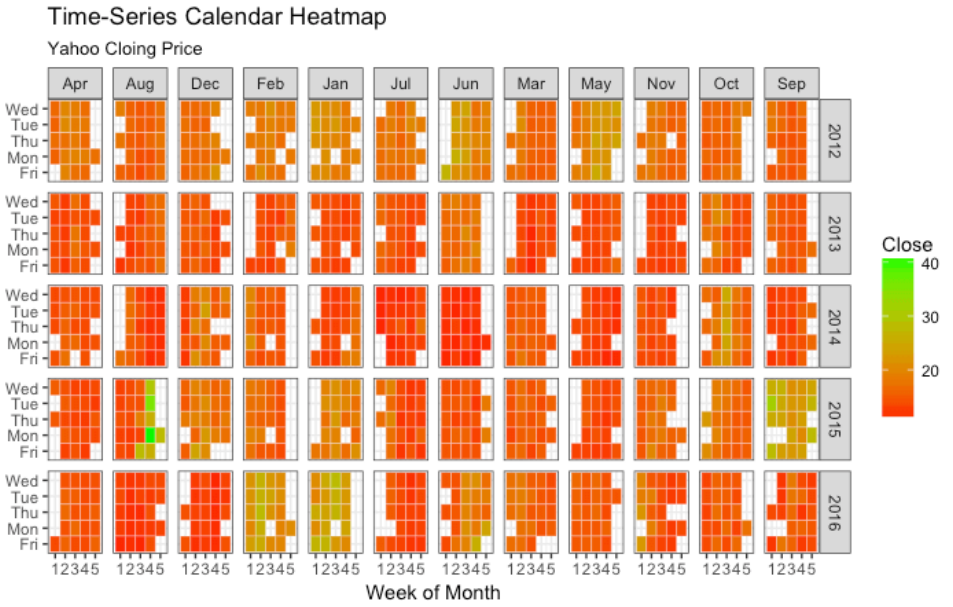


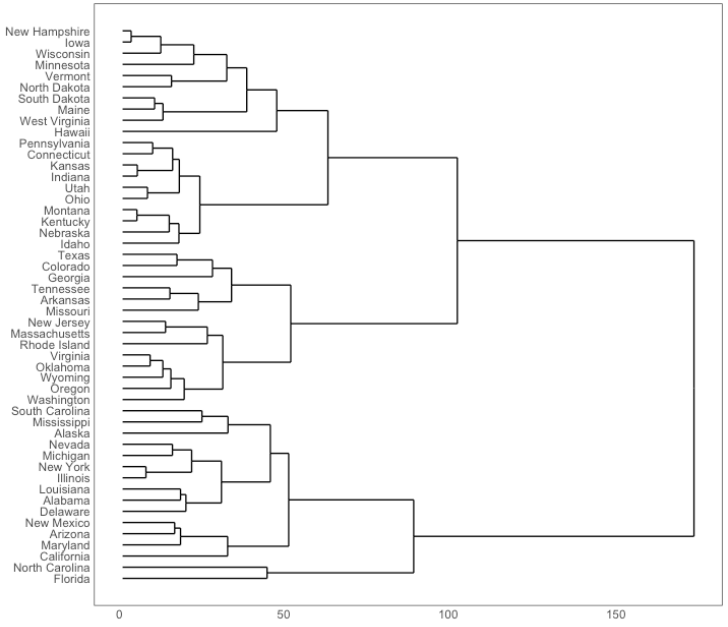
ggplot

Can make all kinds of interesting plots, with minimal coding









http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html

Base plot vs. ggplot2

Which one do you choose?

- There are pros and cons to each

- Base plot
  - Automatically installs with R download
  - Starts off with basics and then add complexity
  - Has different functions for scatterplot, boxplot, histogram
    - `plot(), hist(), boxplot(), etc.`

- ggplot2
  - Can be easier to make more complex graphs
  - All plots are within ggplot framework
  - Lots of help online
  - Constantly updating, can be buggy

https://frinkiac.com

Take home messages

Things to keep in mind

- A good picture of your data
  - May help identify appropriate statistical methods
  - May help identify errors or irregularities

- A really good picture of your data
  - Can tell your story for you
  - Doesn’t have to be complicated

BCC: Biostatistics Collaboration Center

Contact Us

- Request an Appointment
  - http://www.feinberg.northwestern.edu/sites/bcc/contact-us/request-form.html
- General Inquiries
  - bcc@northwestern.edu
  - 312.503.2288
- Visit Our Website
  - http://www.feinberg.northwestern.edu/sites/bcc/index.html

Biostatistics Collaboration Center |680 N. Lake Shore Drive, Suite 1400 |Chicago, IL 60611

Northwestern Medicine®
Feinberg School of Medicine

Your feedback is important to us! (And helps us plan future lectures).

Complete the evaluation survey to be entered in to a drawing to win 2 free hours of biostatistics consultation.