



Introduction to Visualization



Revolution Analytics



1 Agenda

2 Why visualize?

3 Why R?





Outline

1 Agenda

2 Why visualize?

3 Why R?





Overview

- Preliminaries
- R Base Graphics
- Lattice Graphics
- ggplot2 Graphics
- Polishing with themes and the grid package.





Preliminaries

- Functions and packages
- Loading and getting basic info
- Summaries and aggregations
- Saving data, files, and the workspace





Base R Graphics

- Intro
- Plot functions
- Customization





Lattice

- Intro
- Trellis formulas
- Plot functions





ggplot2

- Intro
- Plot functions





Polishing your ggplots

- Intro
- Aesthetics arguments
- Custom themes
- Logos and the grid package





Outline

1 Agenda

2 Why visualize?

3 Why R?





Why visualization?

“The ability to take data-to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it-that’s going to be a hugely important skill in the next decades”

“Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.”

– Hal Varian, Google’s Chief Economist

– The McKinsey Quarterly, Jan 2009





Why data visualization matters

- Helps understand underlying patterns in the data
 - Trends, interactions, and relationships
- More informative than just numbers and words
 - Clarity on what the numbers and words mean
- Aids in the presentation of data
 - “A picture is worth a thousand words”
- Human comprehension
 - The data visualization tells a story





Illustration - What's the difference ?

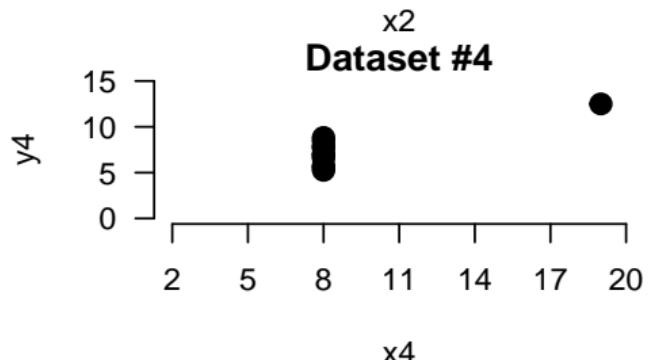
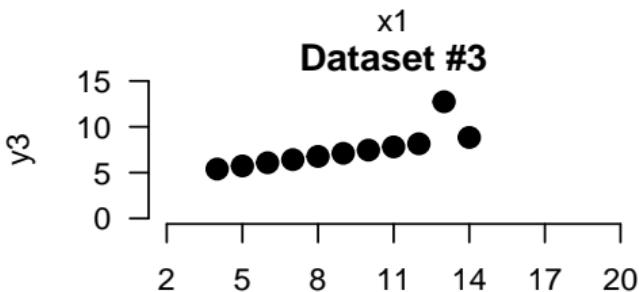
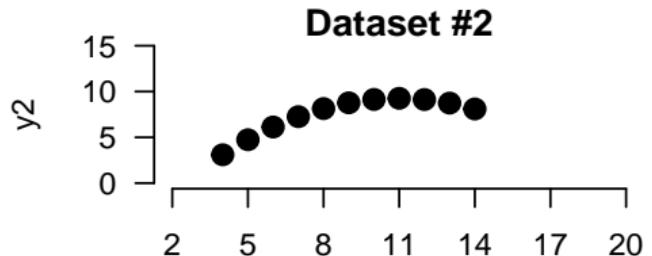
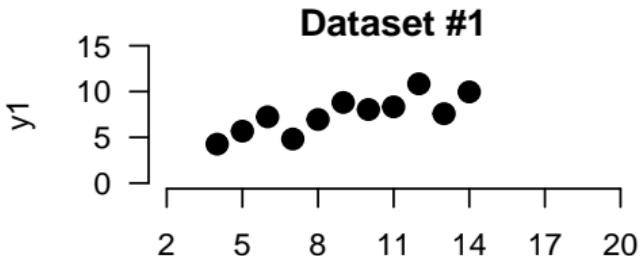
- 4 x-y data sets (Anscombe's quartet)
- Identical means and standard deviation

x1	y1	x2	y2	x3	y3	x4	y4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89





How about this?



{

}





Outline

1 Agenda

2 Why visualize?

3 Why R?





Why visualize with R?

Visualization is a very important component of R

■ CRAN Graphics Task View

Why does the New York Times use R for visualisation?

“R makes it easy to read data, generate lines and points, and place them where you want them. It’s very flexible and super quick. When you’ve only got two or three hours until deadline, R can be brilliant.”

– Amanda Cox, *Graphics Editor*

– blog.revolutionanalytics.com
REVOLUTION
ANALYTICS



Visualizing Friendships on Facebook



<https://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919>





London bike plot



Image produced by James Cheshire at UCL:

<http://spatial.ly/2012/02/great-maps-ggplot2/>





The different plotting paradigms

R has two different graphics paradigms:

base The original graphics engine - pen and paper model

grid Uses viewports to define windows on a grid. Object oriented. Two very popular packages make use of grid: lattice and ggplot2





Grid Packages

lattice The original trellis package to create multiple small subplots.

ggplot2 A very popular way of building plots from components, based on the theory of *grammar of graphics*





Easy integration with other systems

- R + GoogleVis
- R + Tableau
- R + JavaScript
- etc...

Let's get started!





Thank you

Revolution Analytics is the leading commercial provider of software and support for the popular open source R statistics language.



www.revolutionanalytics.com

1.855.GET.REVO

Twitter: @RevolutionR