Tools!

Jackie Wirz & Steven Bedrick
CONJ 610 12/17/2014

Plan for today:

Discussion of types of tools

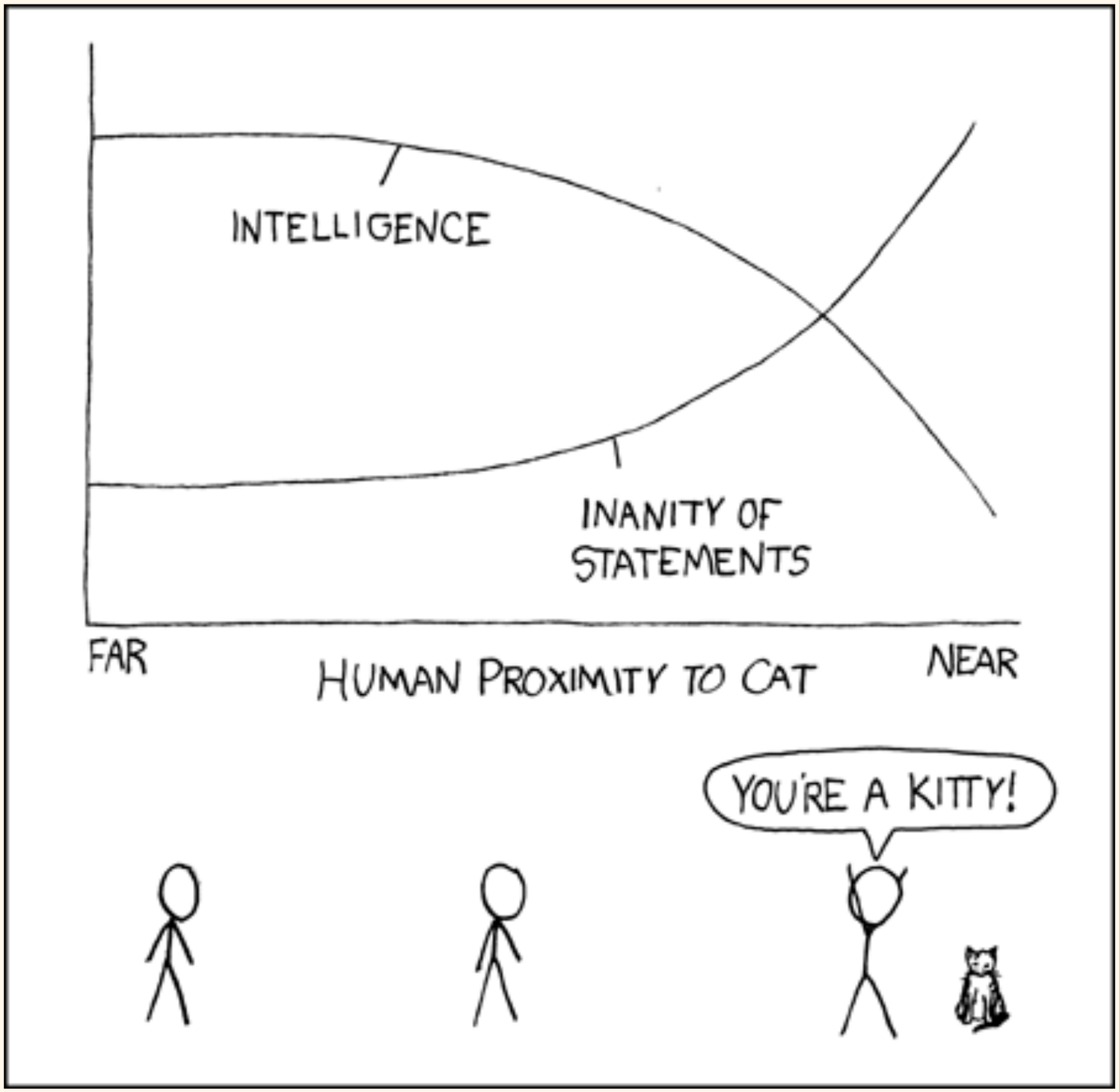
Example tools

Workflow considerations

Other considerations

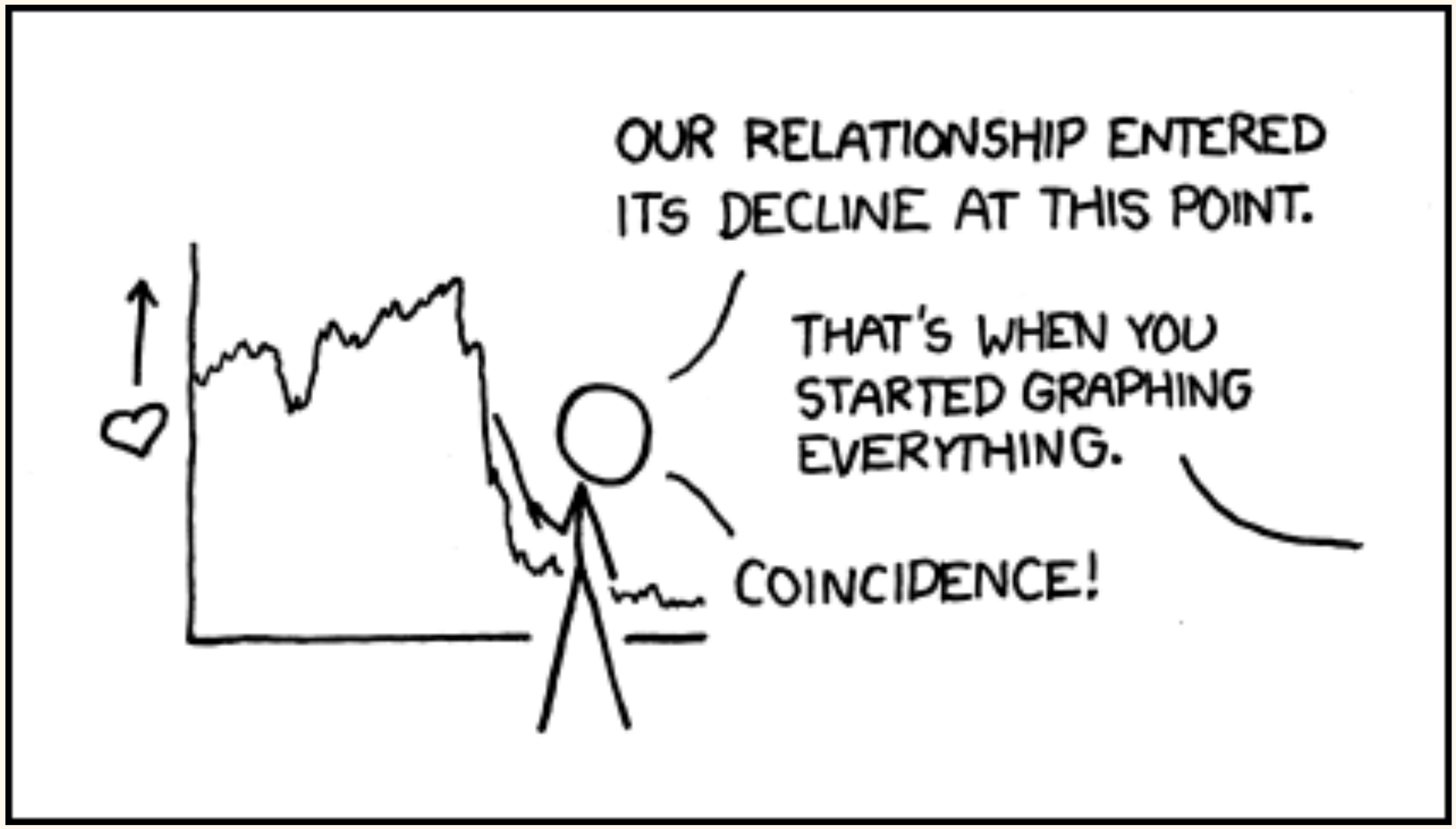
Avoid the default settings!

First: It is possible to do good visualization without a computer!

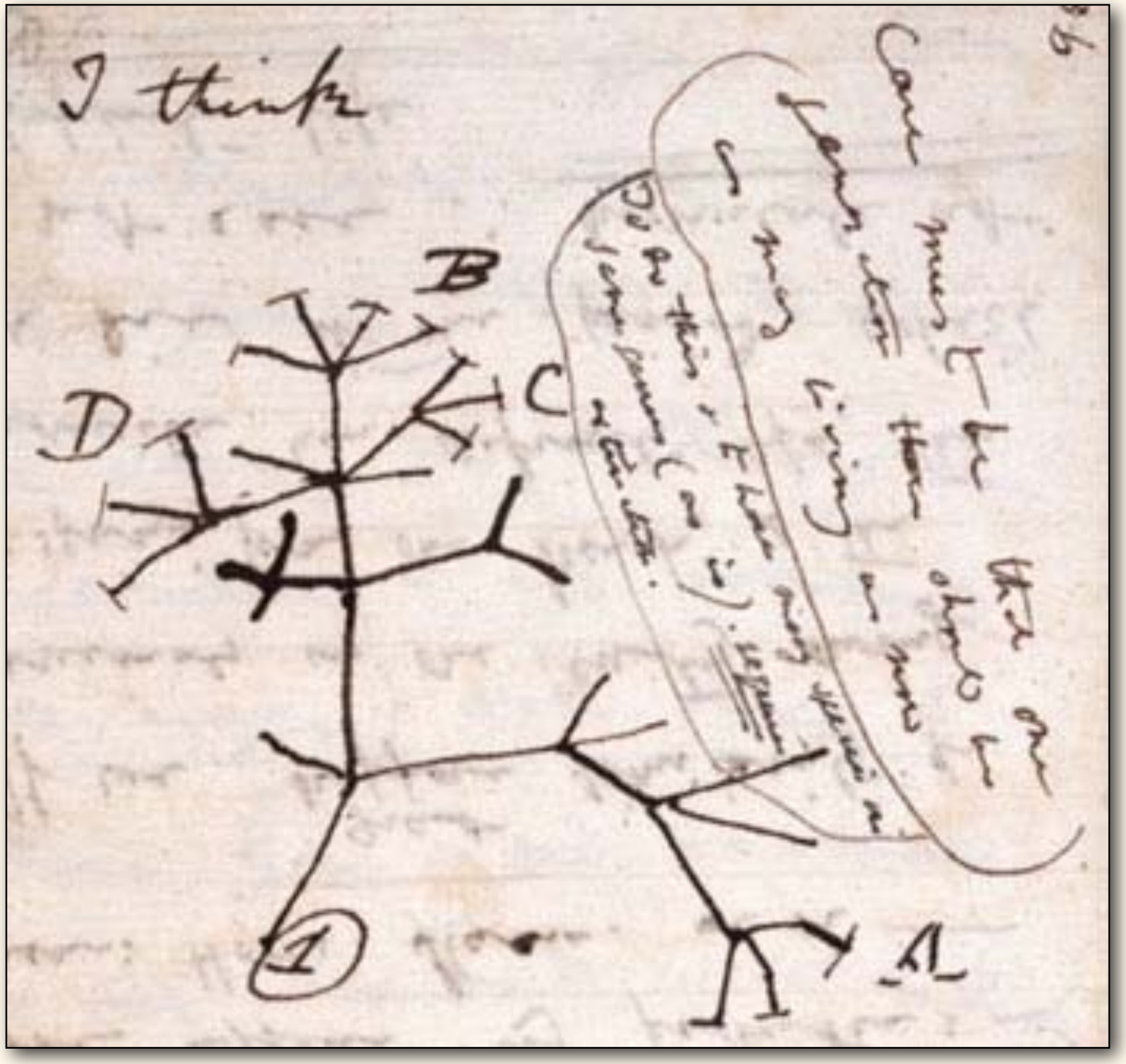


http://xkcd.com/231/

First: It is possible to do good visualization without a computer!



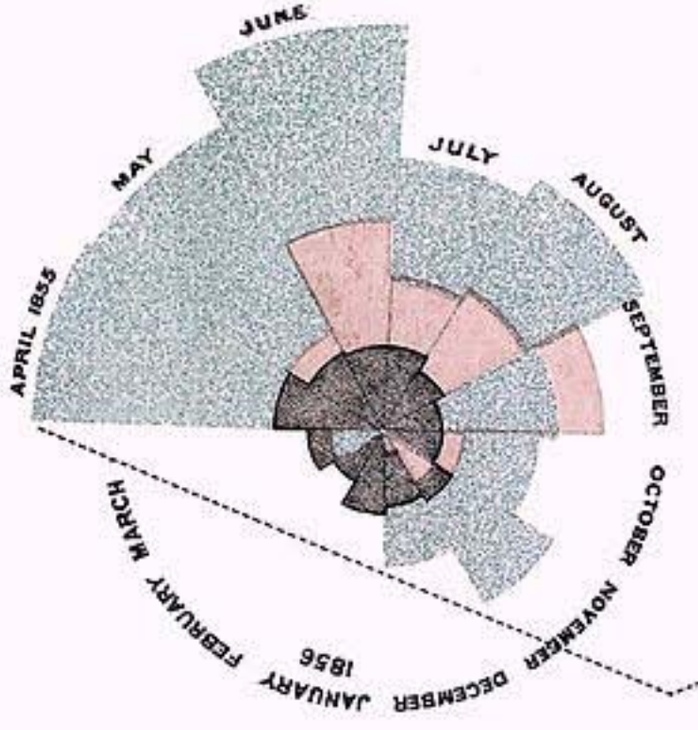
http://xkcd.com/523/



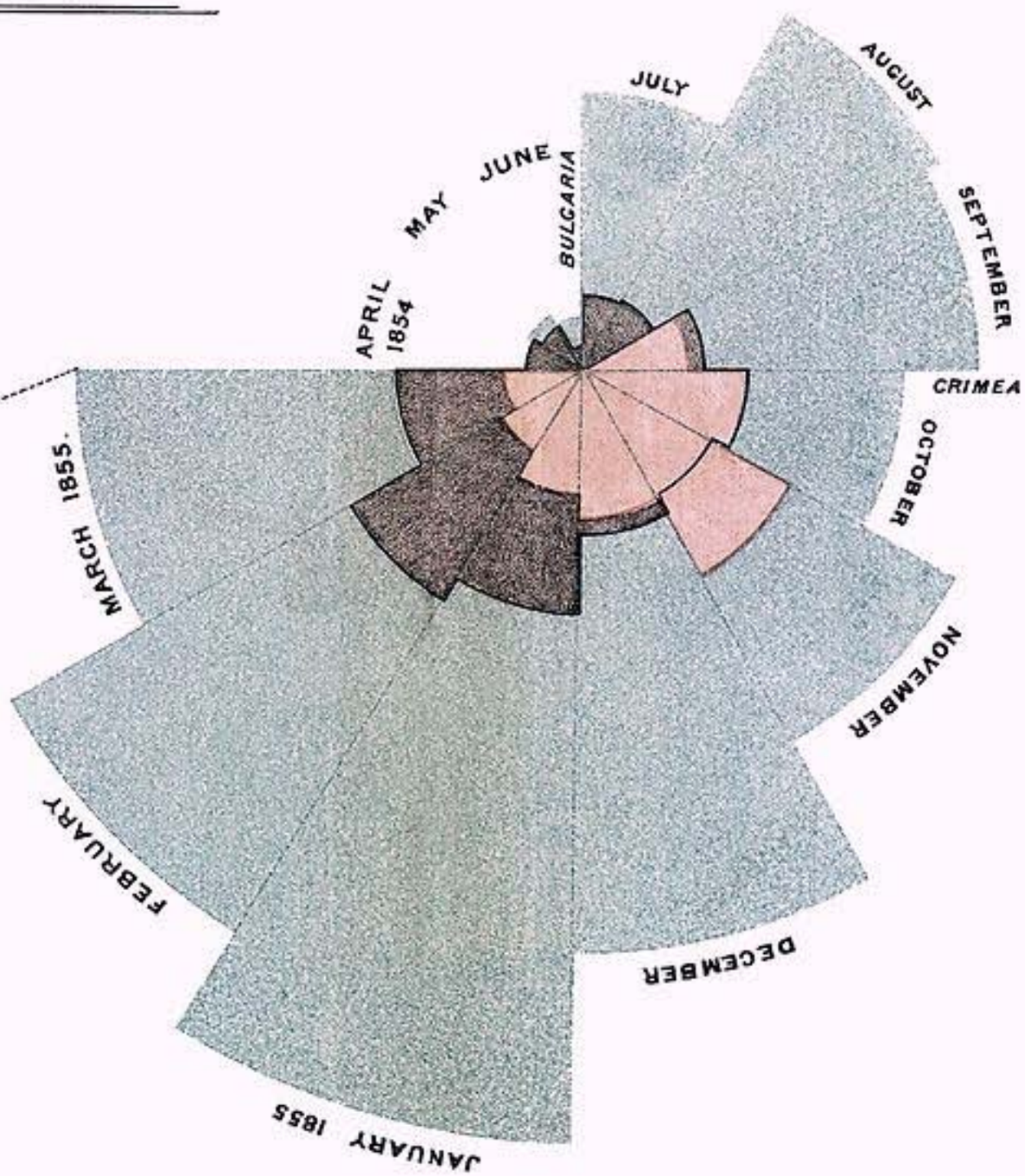
http://www.nhm.ac.uk/nature-online/evolution/tree-of-life/darwin-tree/

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

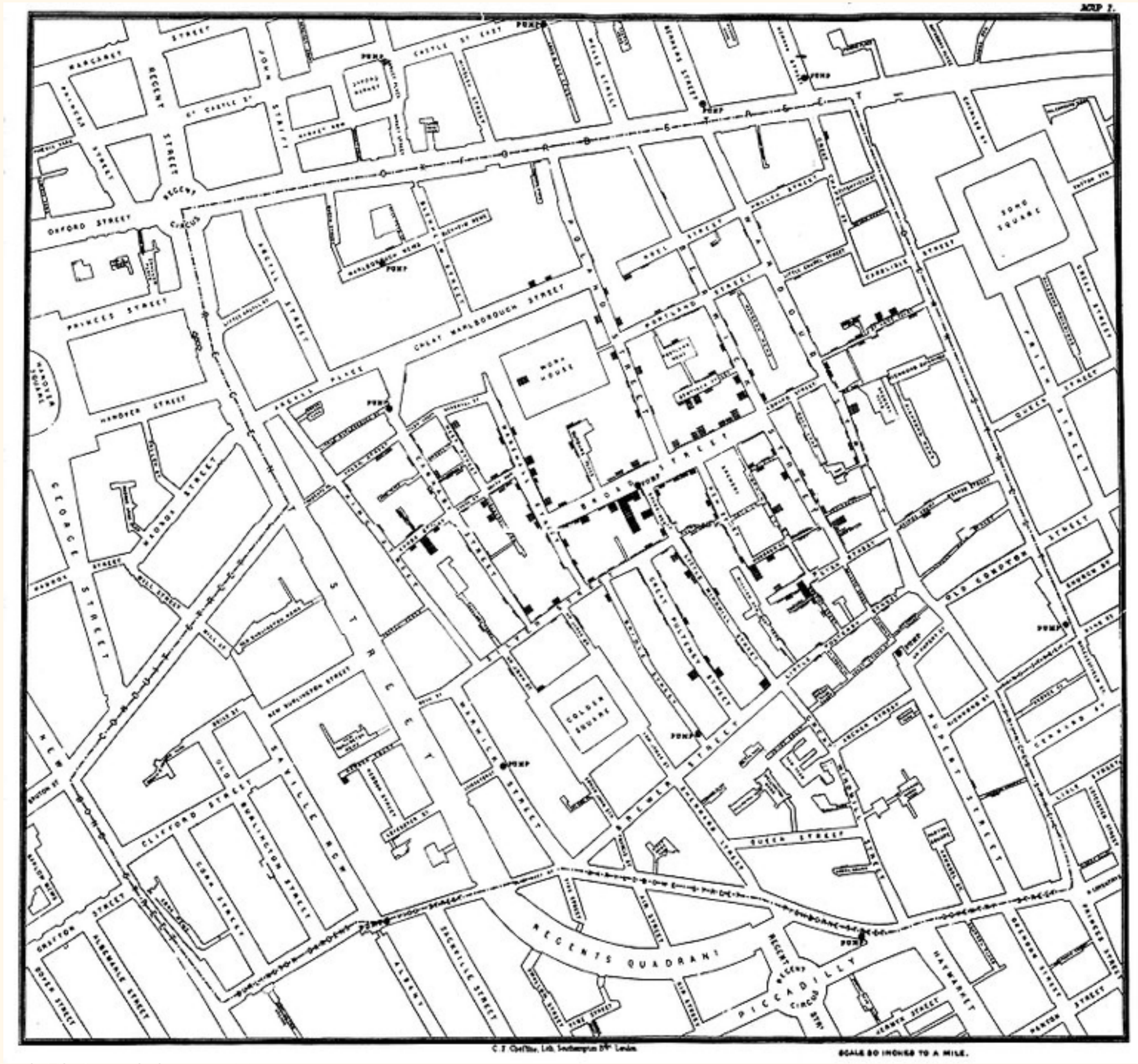


1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

http://en.wikipedia.org/wiki/File:Nightingale-mortality.jpg



http://en.wikipedia.org/wiki/File:Snow-cholera-map-1.jpg

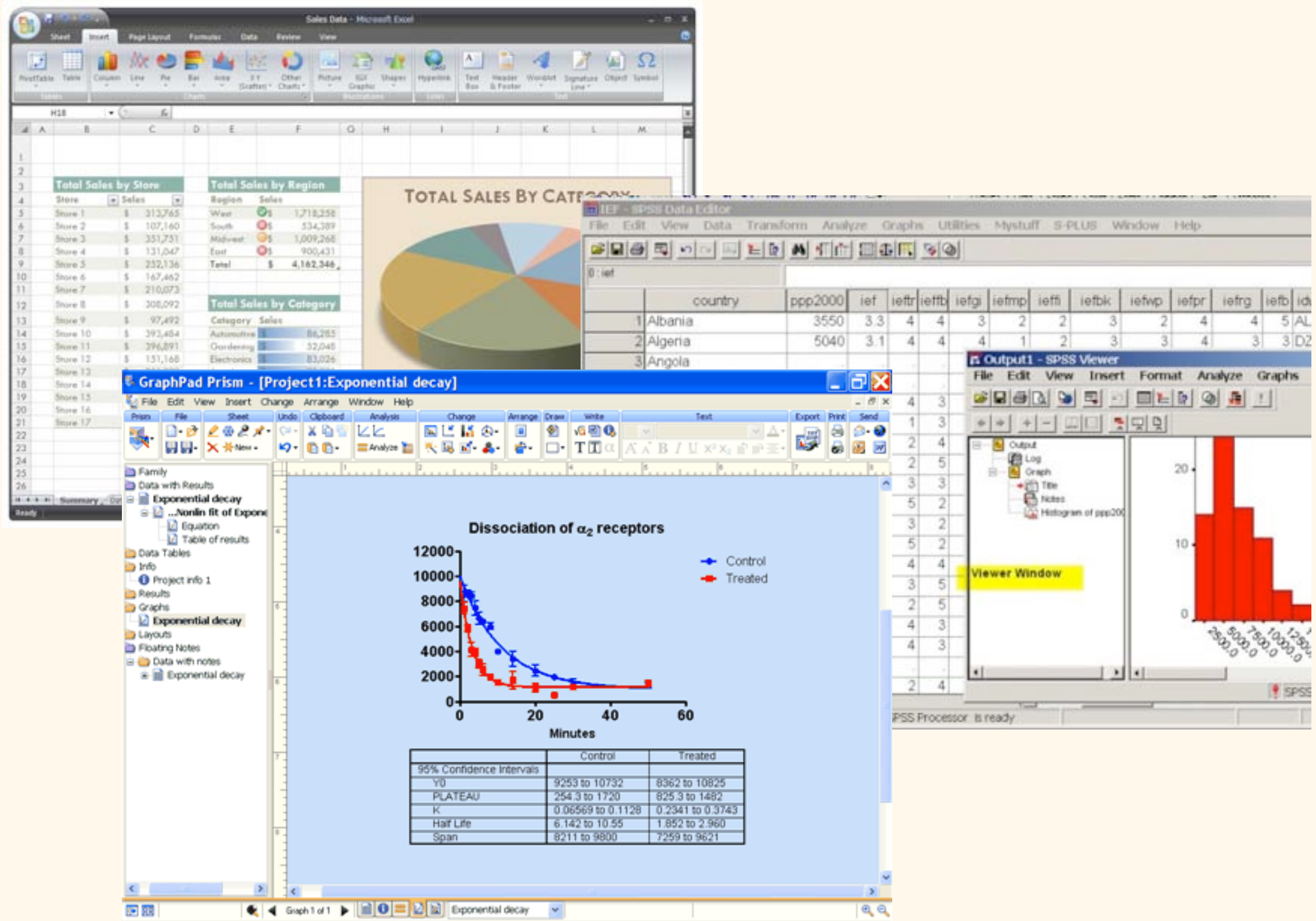
That said, computers make it a lot easier.

There are two main categories of tool:

1. Data analysis software that also does visualization;

2. Dedicated visualization packages.

1. Data analysis software that also does visualization;



coffee.xlsx

125%

Search in Sheet

Home | Layout | Tables | Charts | Chart Layout | Format | SmartArt | Formulas | Data | Review

Change Chart Type: Column, Line, Pie, Bar, Area, Scatter, Other

Insert Sparklines: Line, Column, Win/Loss

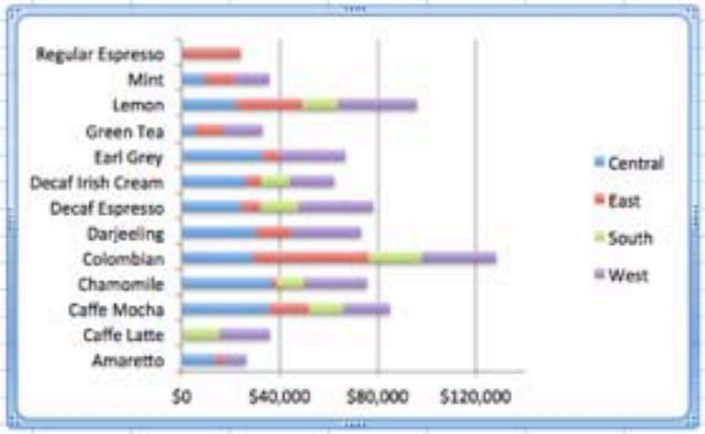
Data: Select, Switch Plot

Chart Quick Layouts

Chart Styles

Chart 2

	A	B	C	D	E
1		Market			
2	Product	Central	East	South	West
3	Amaretto	$14,012	$2,994		$9,263
4	Caffe Latte			$15,443	$20,456
5	Caffe Mocha	$35,218	$18,846	$14,168	$18,874
6	Chamomile	$36,571	$2,193	$11,183	$25,631
7	Colombian	$28,911	$47,385	$21,663	$30,352
8	Darjeeling	$30,284	$14,094		$28,773
9	Decaf Espresso	$24,483	$7,720	$15,381	$30,578
10	Decaf Irish Cream	$26,157	$6,262	$11,596	$18,233
11	Earl Grey	$32,883	$8,507		$27,382
12	Green Tea	$5,209	$11,576		$16,065
13	Lemon	$21,982	$27,177	$14,494	$32,273
14	Mint	$9,335	$11,991		$14,384
15	Regular Espresso		$24,031		



Sheet 1

Normal View | Ready | Sum = 0

Format Data Series

Options
Axis
Order
Error Bars
Fill
Line
Shadow
Glow & Soft Edges
3-D Format

Solid | Gradient | Picture or Texture | Pattern

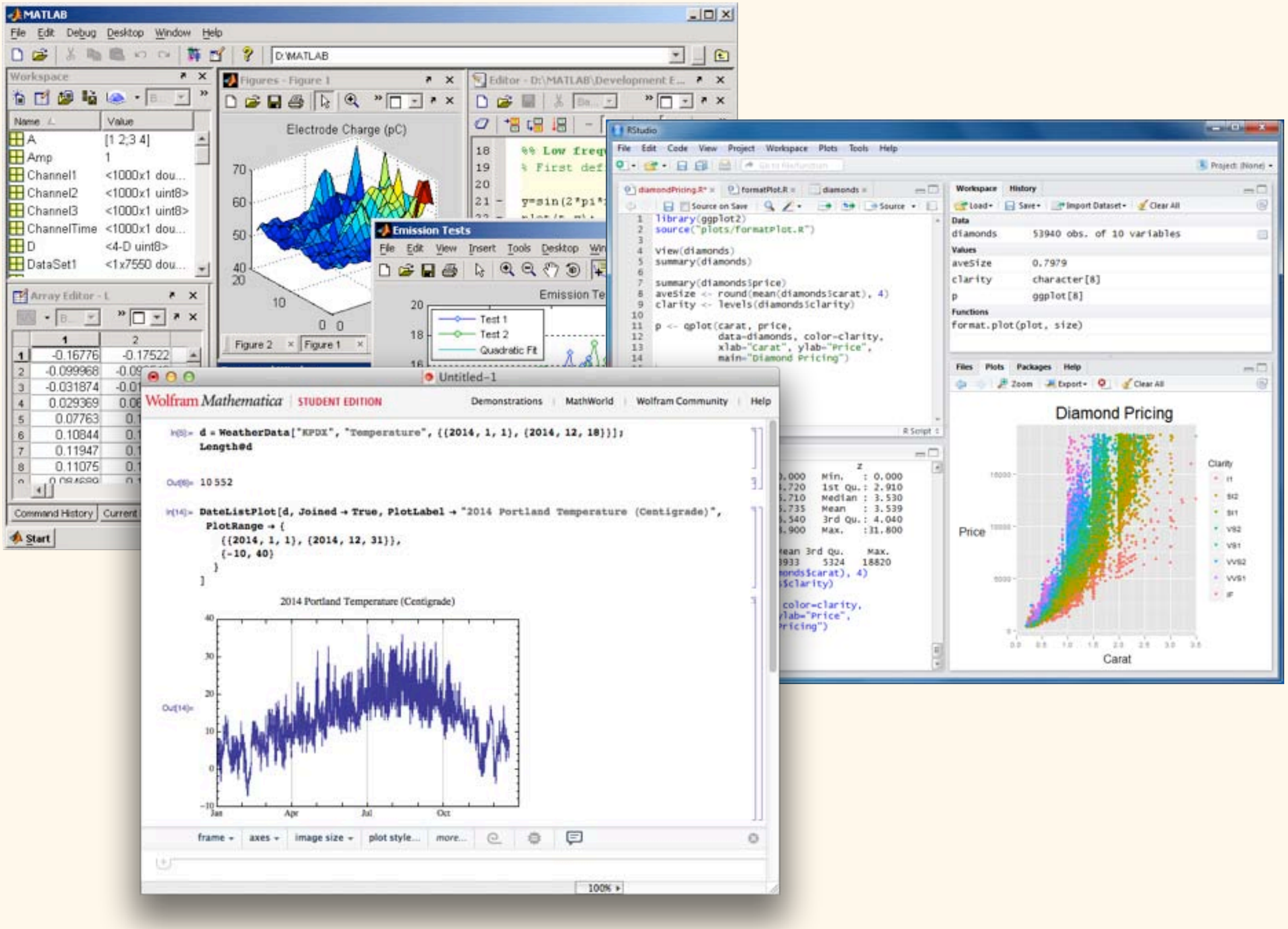
Color: Automatic

Transparency: 0 100

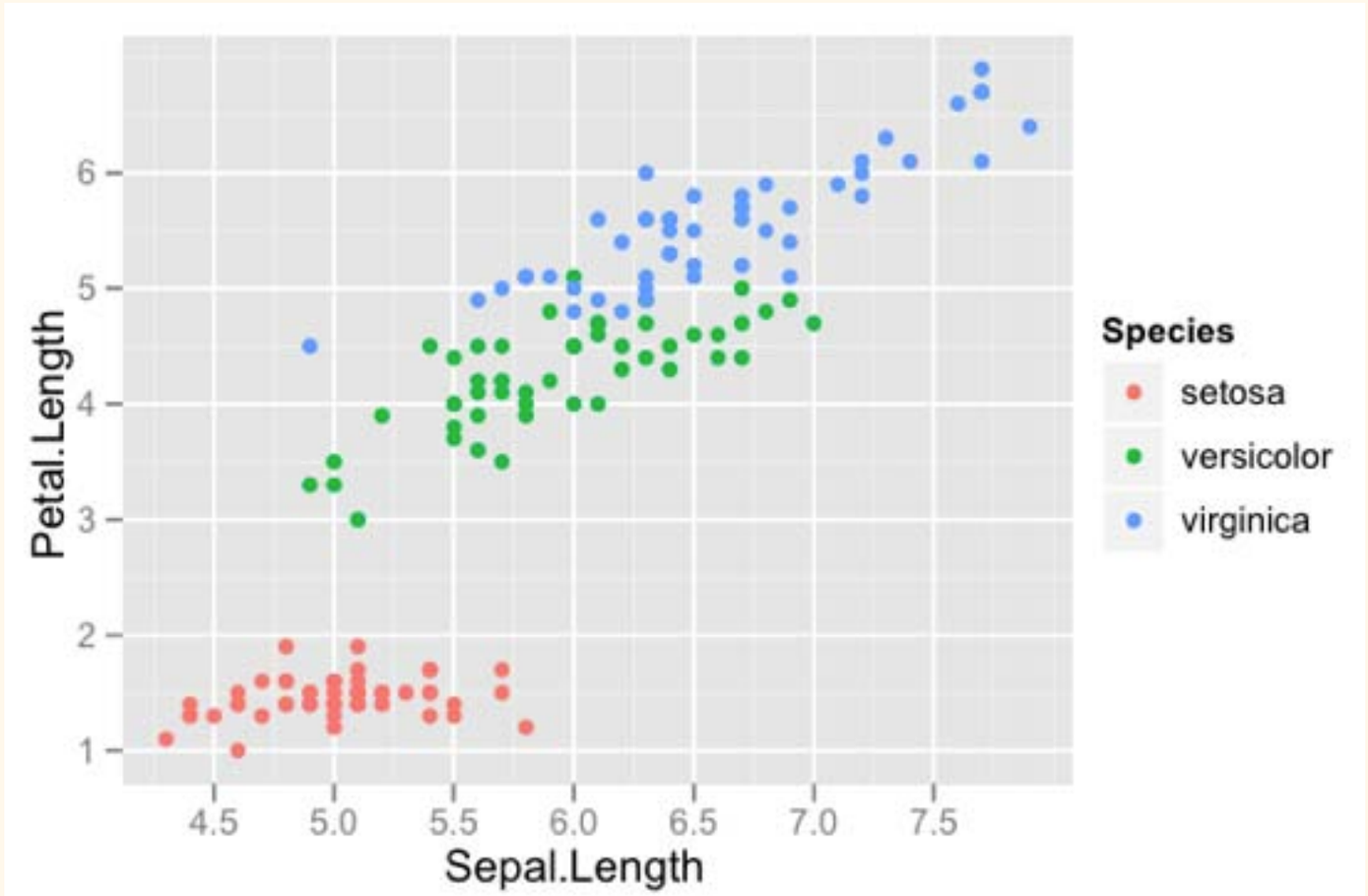
Invert if negative

Cancel | OK

1. Data analysis software that also does visualization;

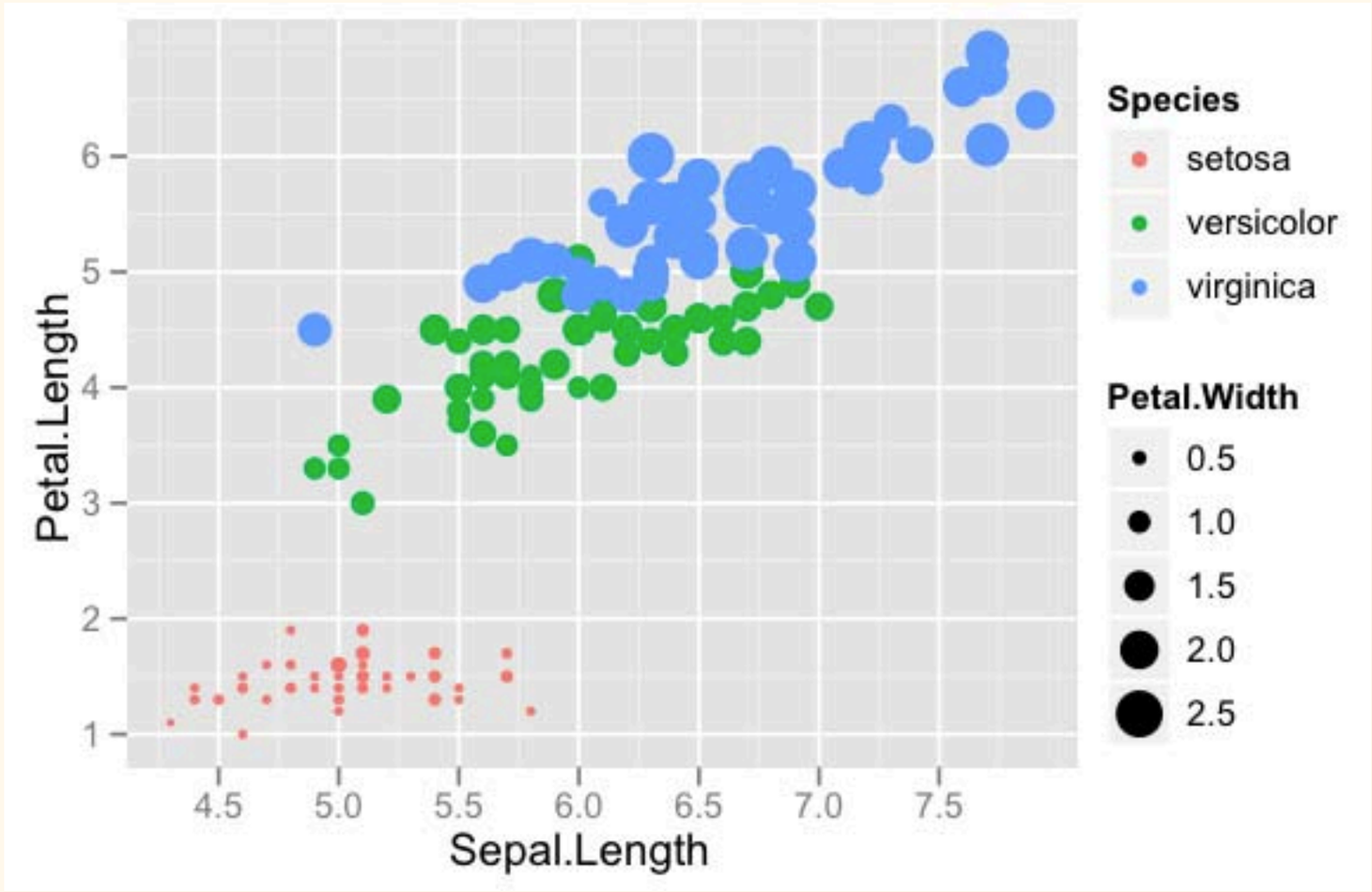


```
qplot(Sepal.Length, Petal.Length, data = iris, color = Species)
```



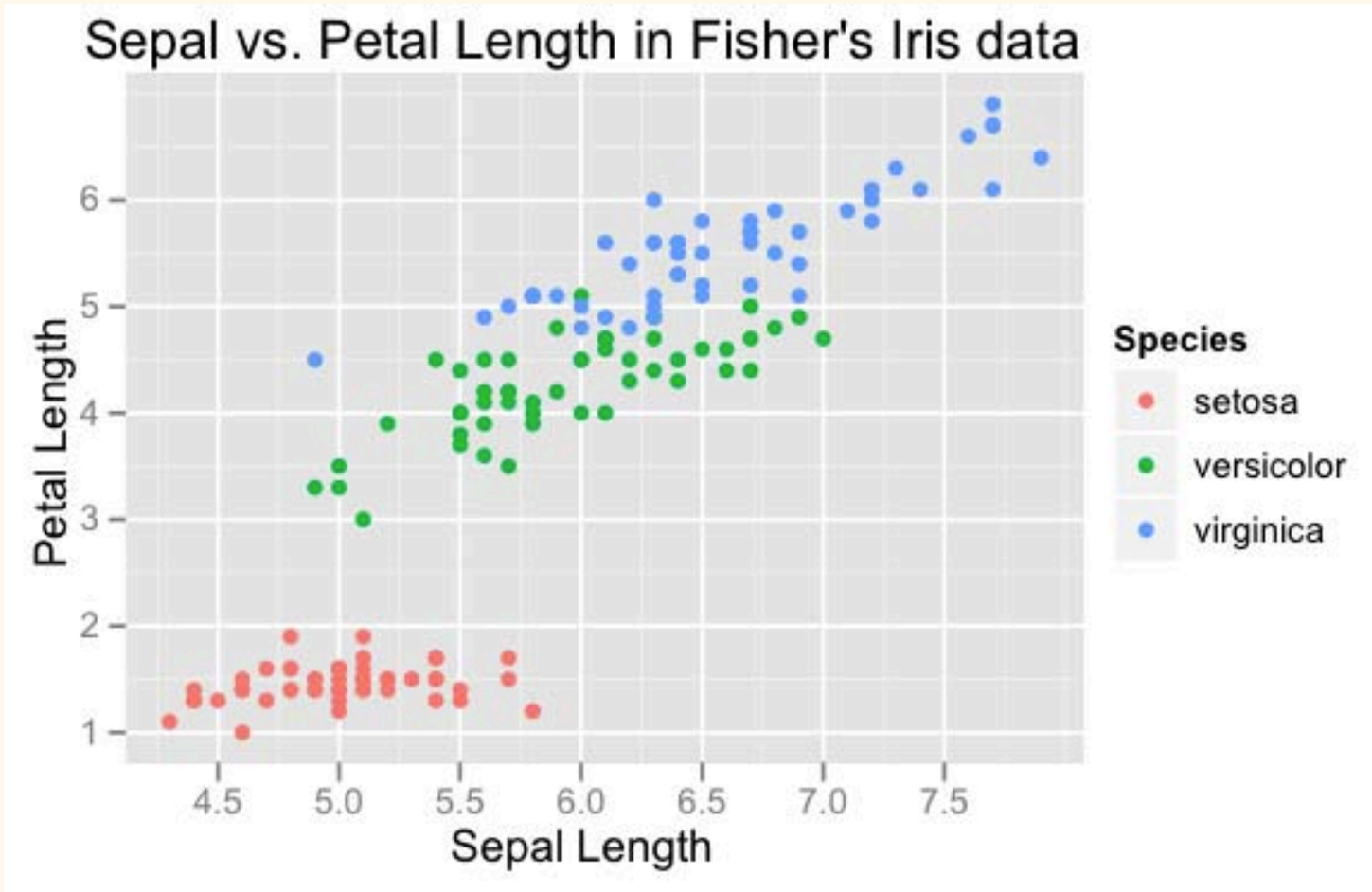
http://blog.echen.me/2012/01/17/quick-introduction-to-ggplot2/

```
qplot(Sepal.Length, Petal.Length, data = iris, color = Species, size = Petal.Width)
```



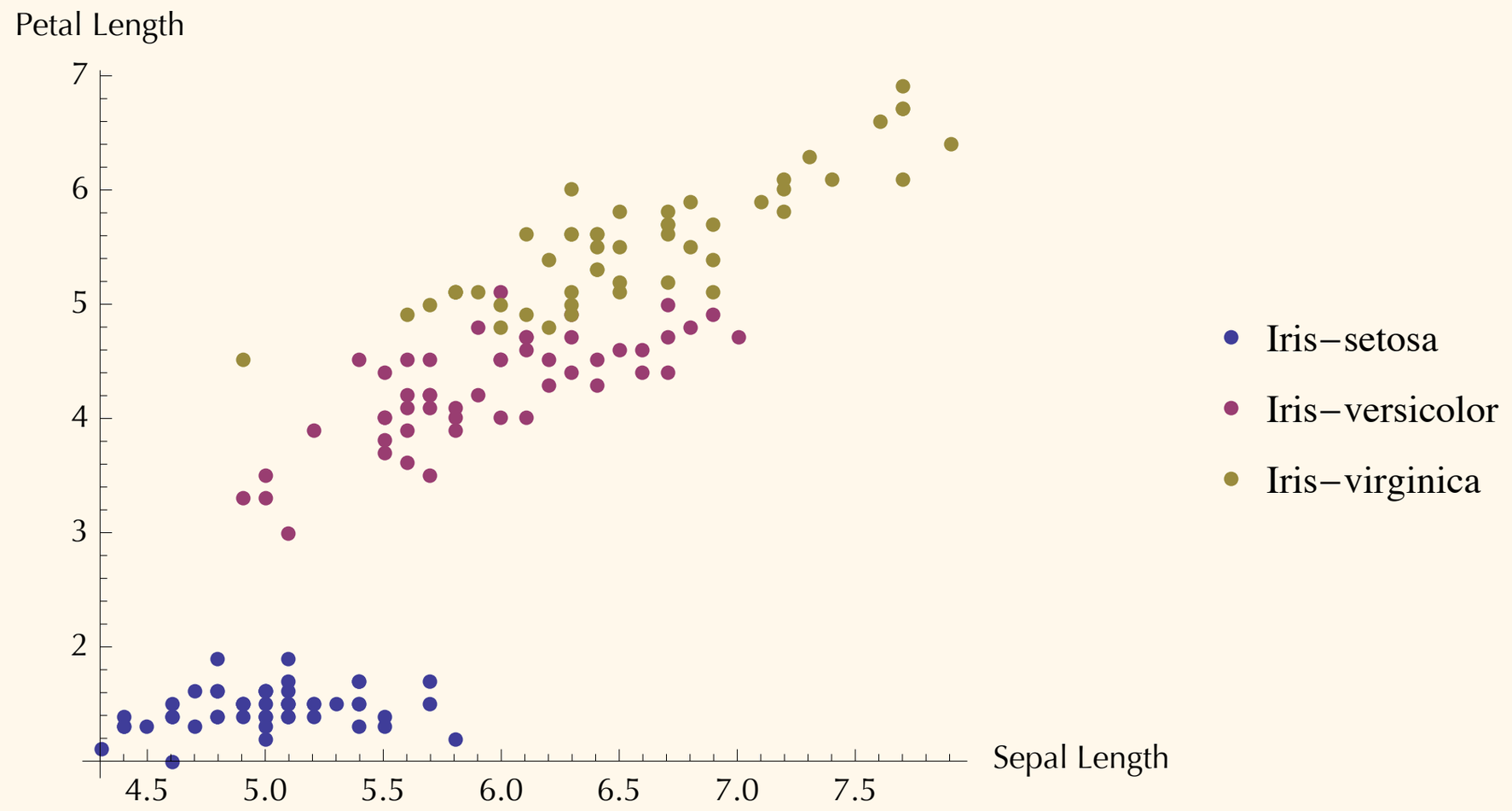
http://blog.echen.me/2012/01/17/quick-introduction-to-ggplot2/

```
qplot(Sepal.Length, Petal.Length, data = iris, color = Species,
    xlab = "Sepal Length", ylab = "Petal Length",
    main = "Sepal vs. Petal Length in Fisher's Iris data")
```



http://blog.echen.me/2012/01/17/quick-introduction-to-ggplot2/

```
ListPlot[
 irises[[All, All, {1, 3}]],
 PlotRange → All, PlotStyle → PointSize[Medium],
 BaseStyle → {FontFamily → "Optima"},
 PlotLegends → z[[All, 1, 5]],
 AxesLabel → {"Sepal Length", "Petal Length"},
 AspectRatio → 2 / 2.5
]
```



Graphical systems (Excel, etc.)

Pro:

"Easier"

Quick path to results, esp. for simple plots

Might be what you're already using

One-stop-shopping (analysis & vis. together)

Easier collaboration, probably

Graphical systems (Excel, etc.)

Cons:

- Less control
- Hideous/boring defaults
- Lots of repetitive clicking around
- Capabilities are (often) limited
- Hard to exactly reproduce “what you did last time”

“Programatic” systems (R, etc.)

Pro:

“Easier”

More control

Easier reproducibility

R is free!

Fewer limits

Default settings for graphs are less ugly

“Programatic” systems (R, etc.)

Con:

Very steep learning curve

Getting around limitations can be a lot of work

Can be a huge time-sink (especially at first!)

Neutral:

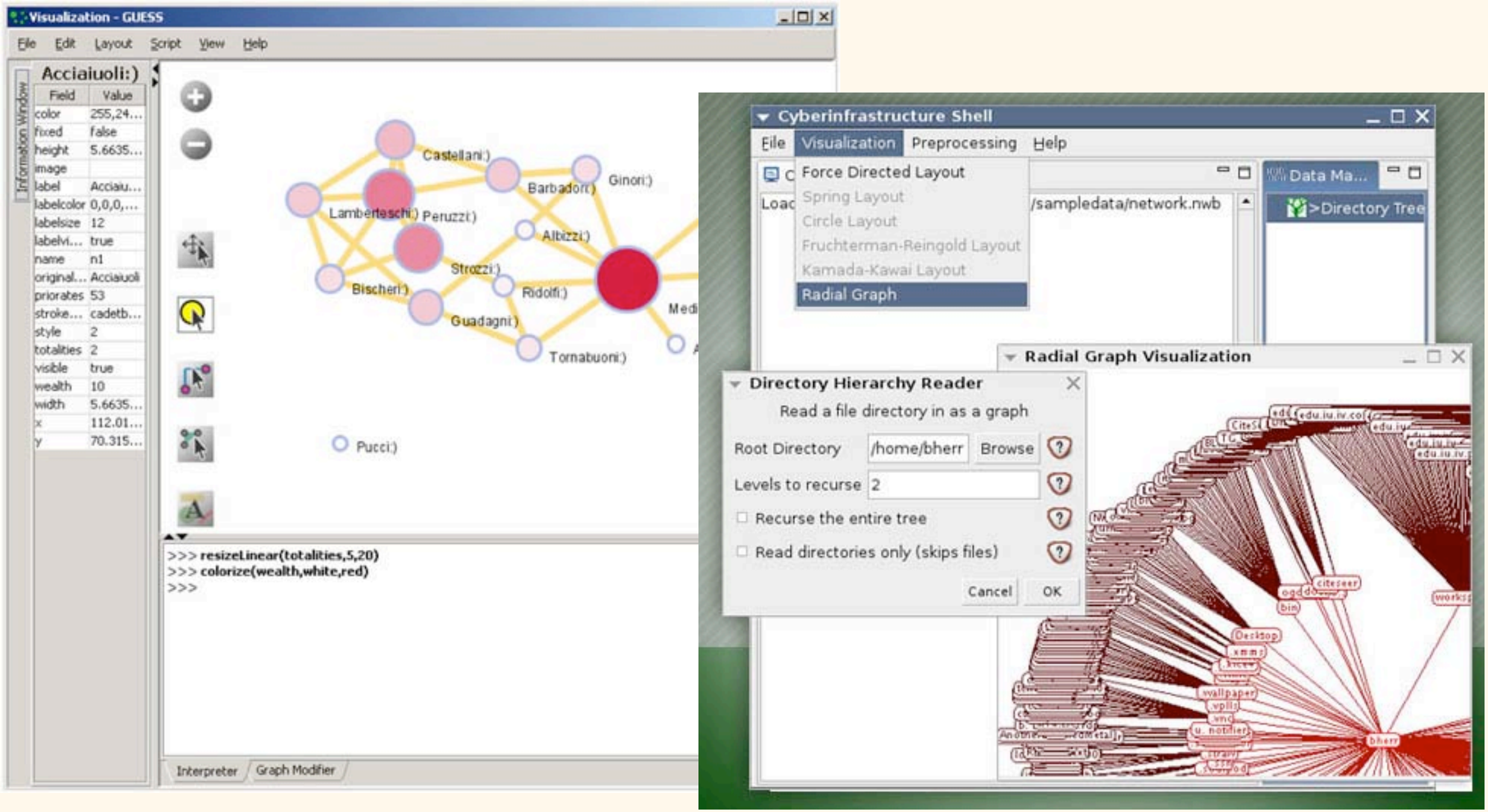
May make collaboration easier or harder, depending.

1. Data analysis software that also does visualization;

2. Dedicated visualization packages.

2. Dedicated visualization packages.

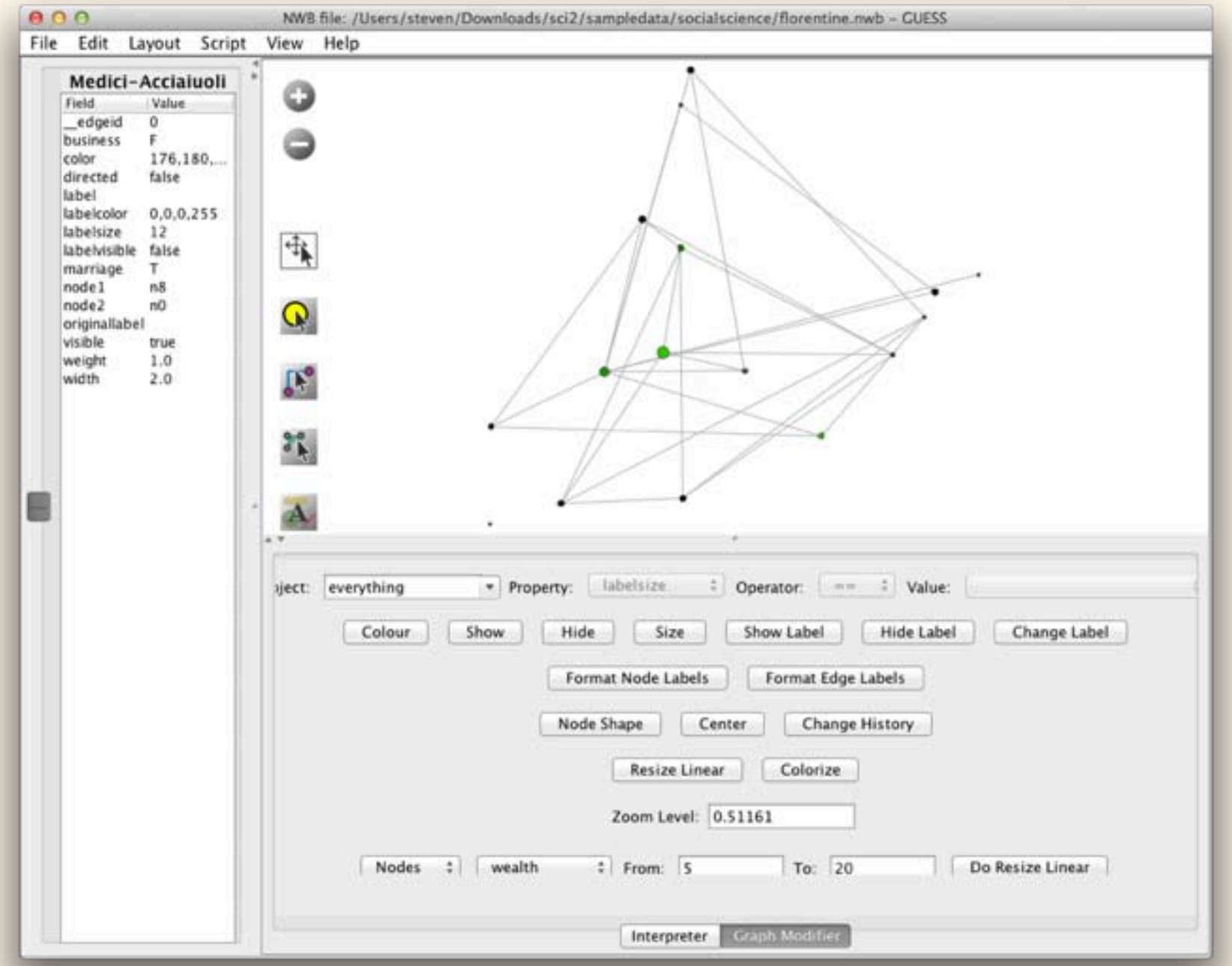
Sci2: network diagrams, simple maps, etc.



https://sci2.cns.iu.edu/user/index.php

2. Dedicated visualization packages.

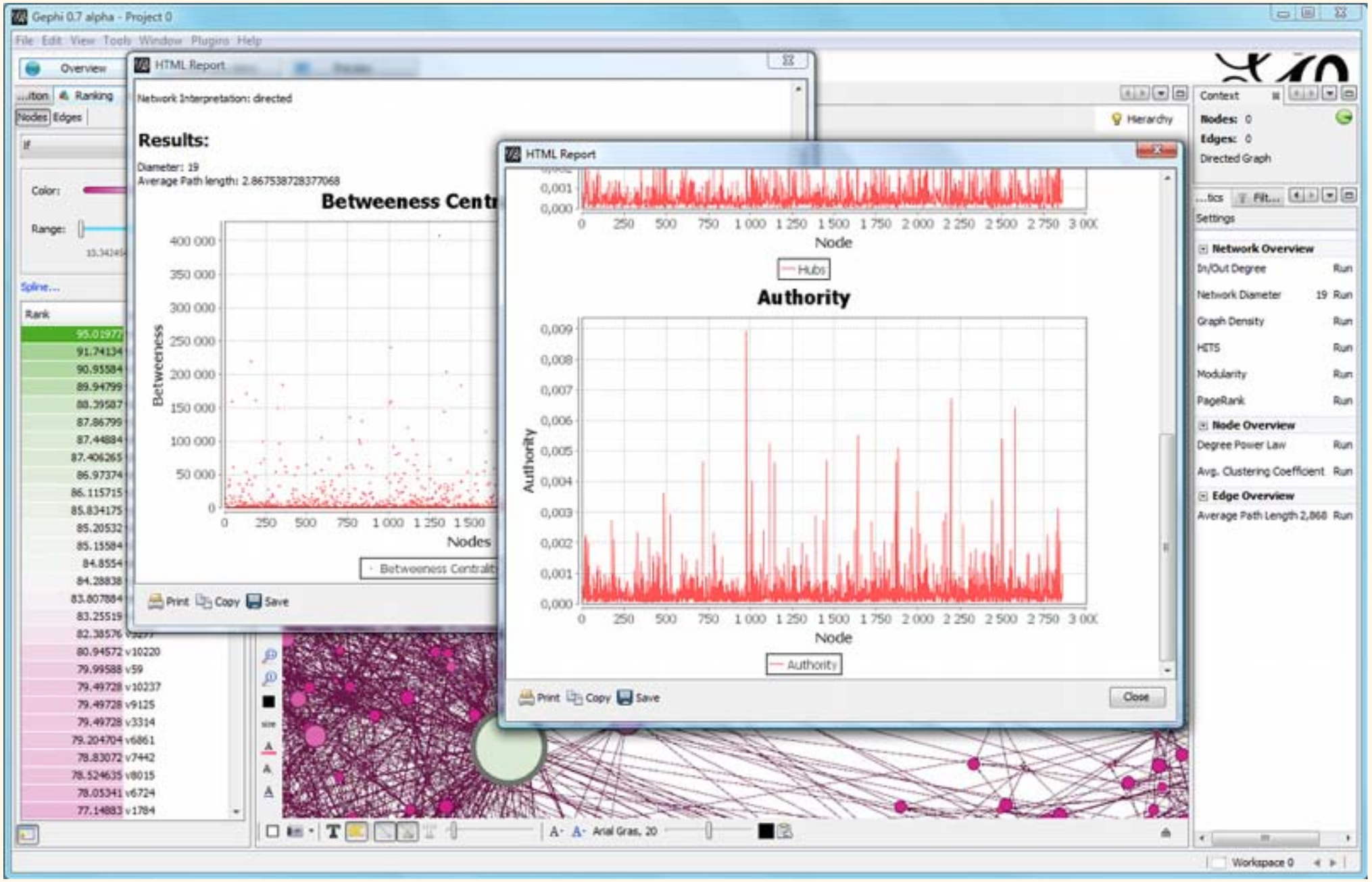
Sci2: network diagrams, simple maps, etc.



https://sci2.cns.iu.edu/user/index.php

2. Dedicated visualization packages.

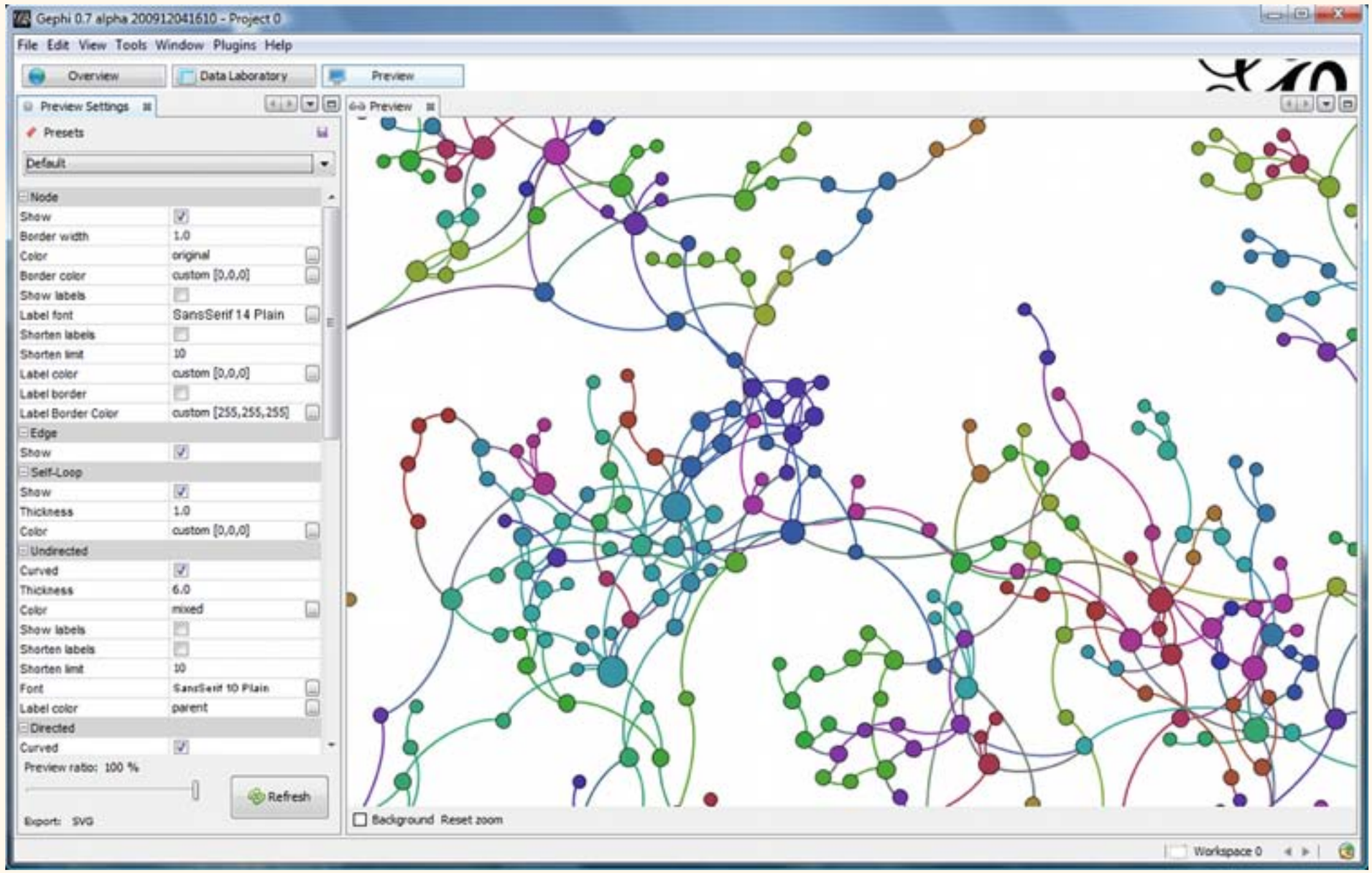
Gephi: network diagrams



https://gephi.github.io

2. Dedicated visualization packages.

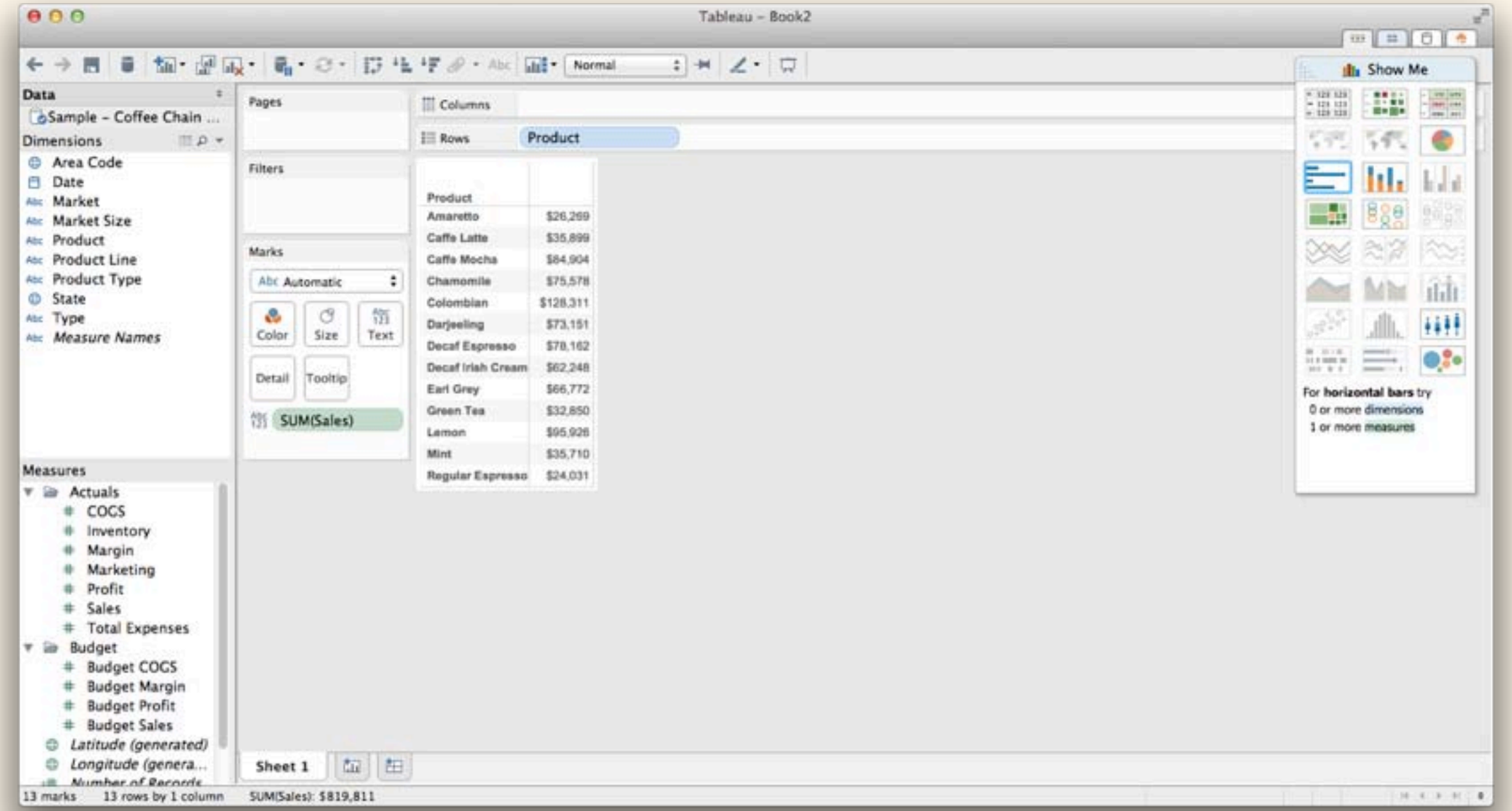
Gephi: network diagrams



https://gephi.github.io

2. Dedicated visualization packages.

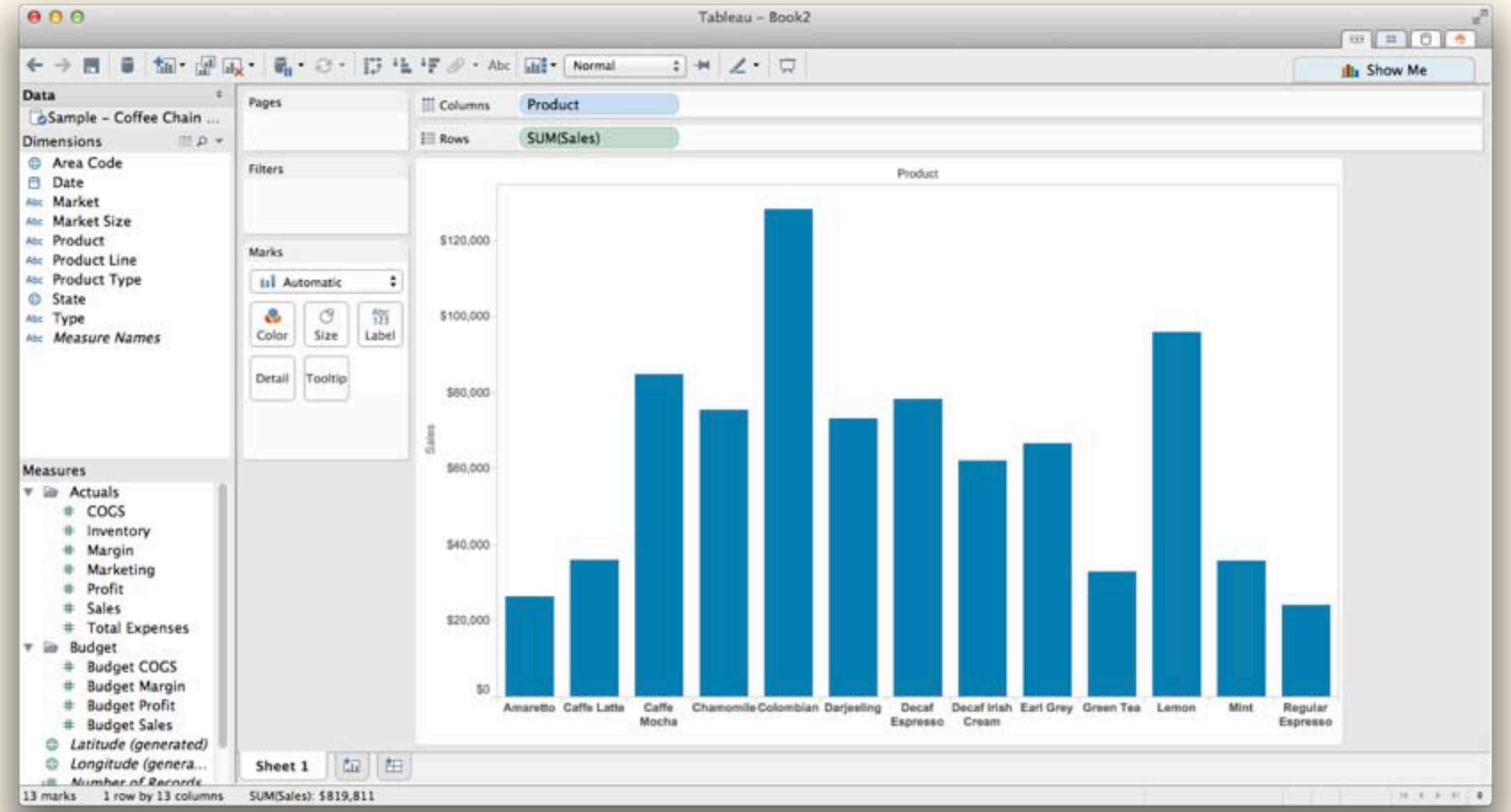
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

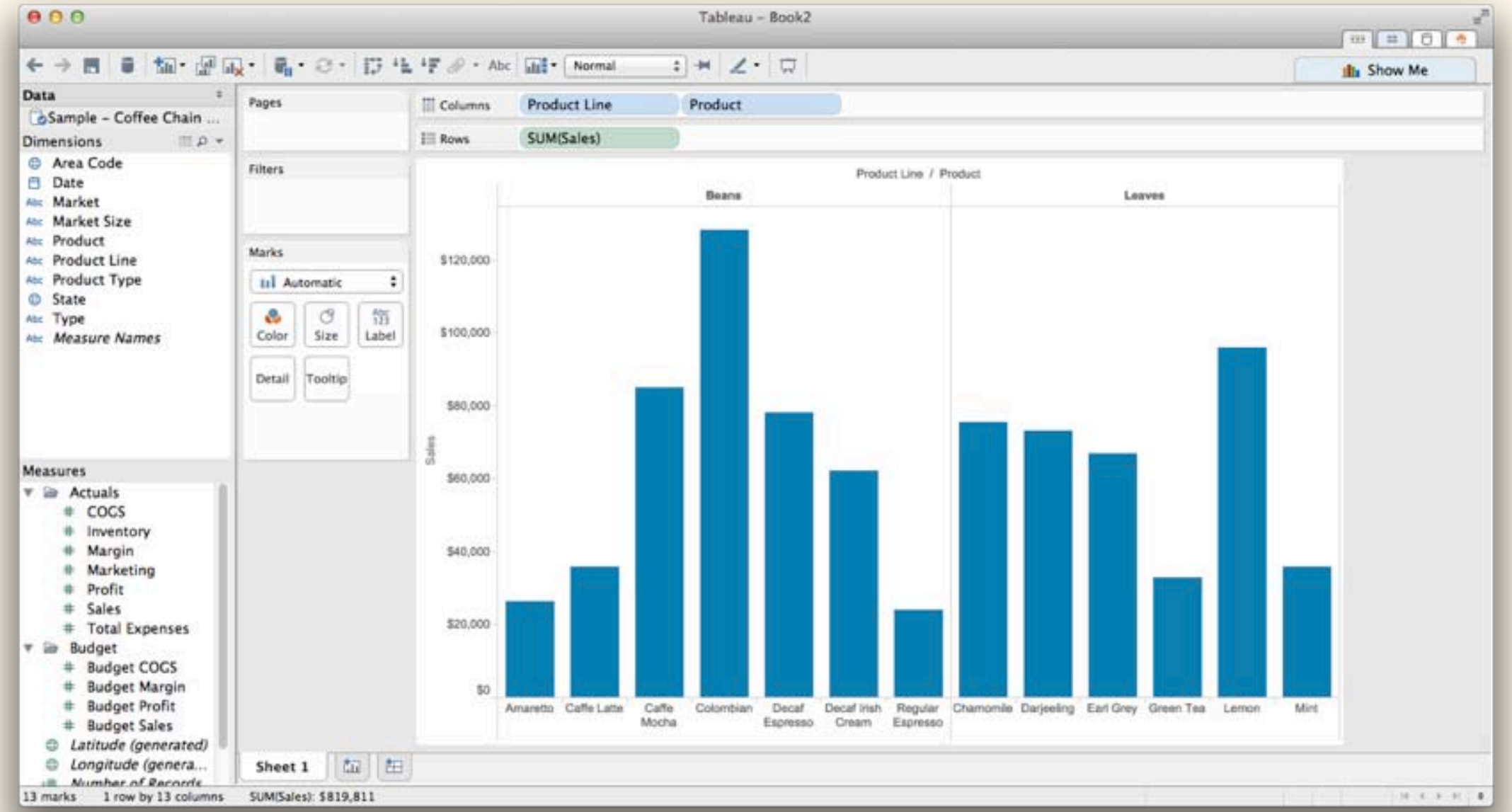
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

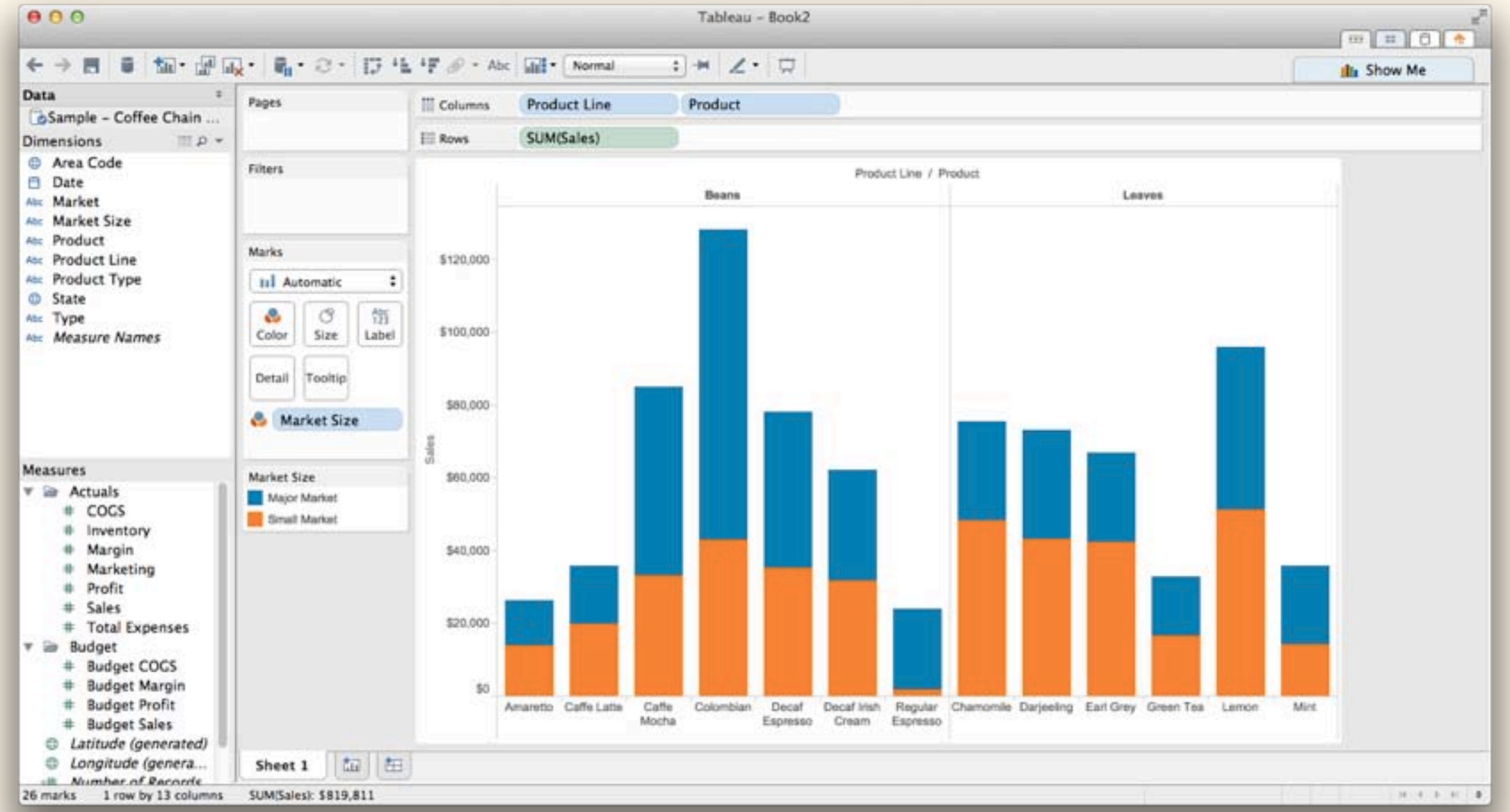
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

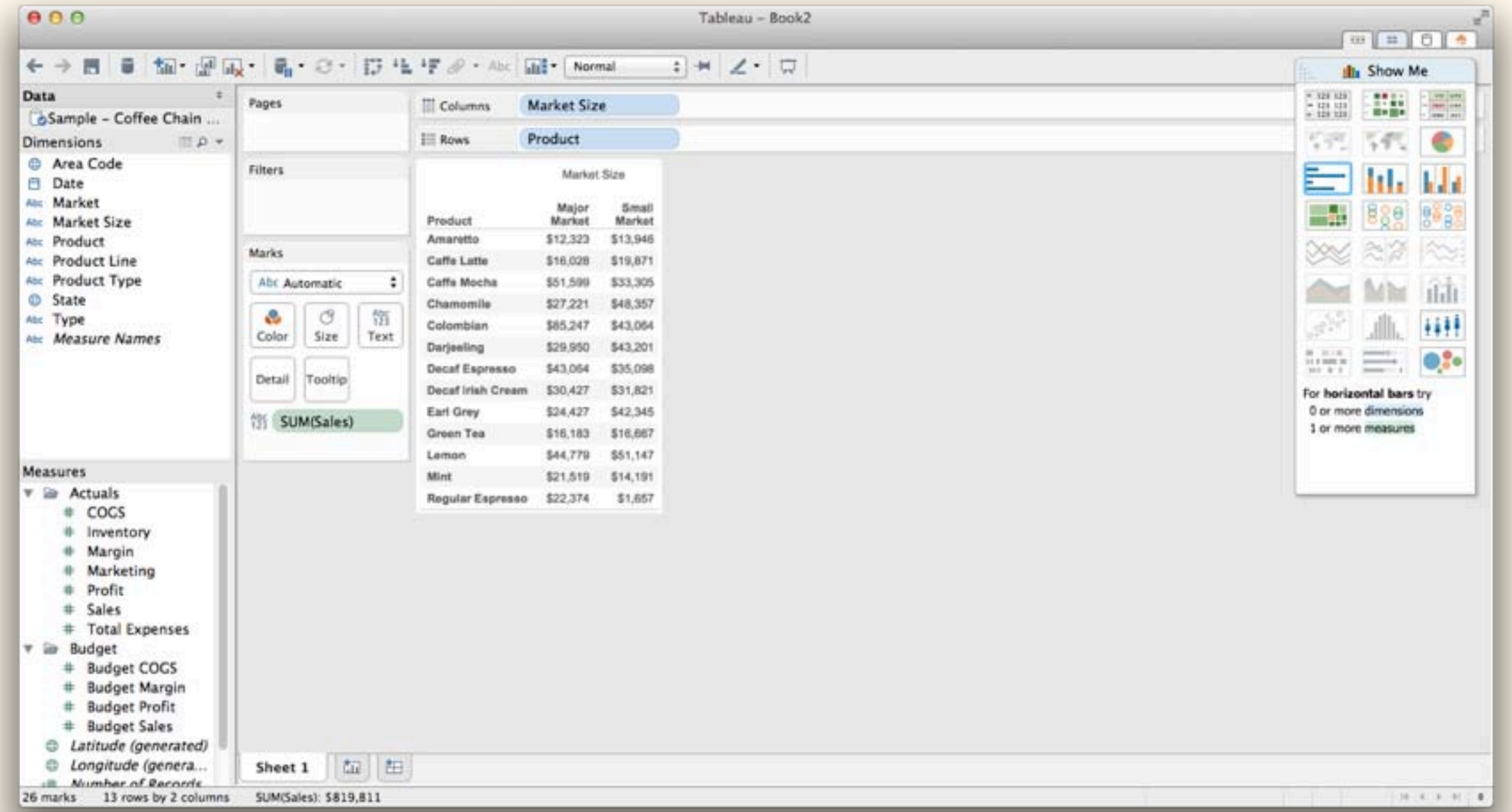
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

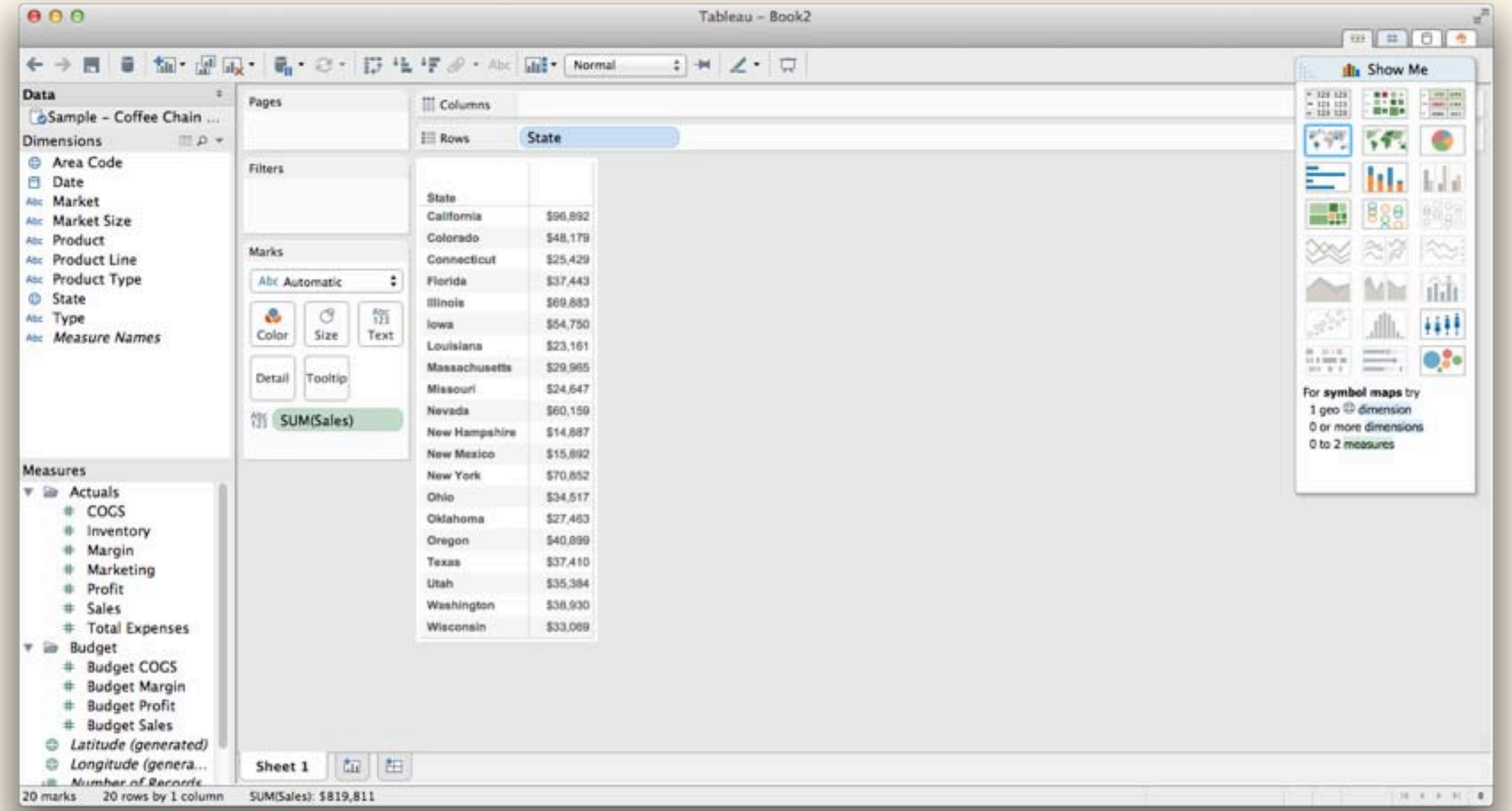
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

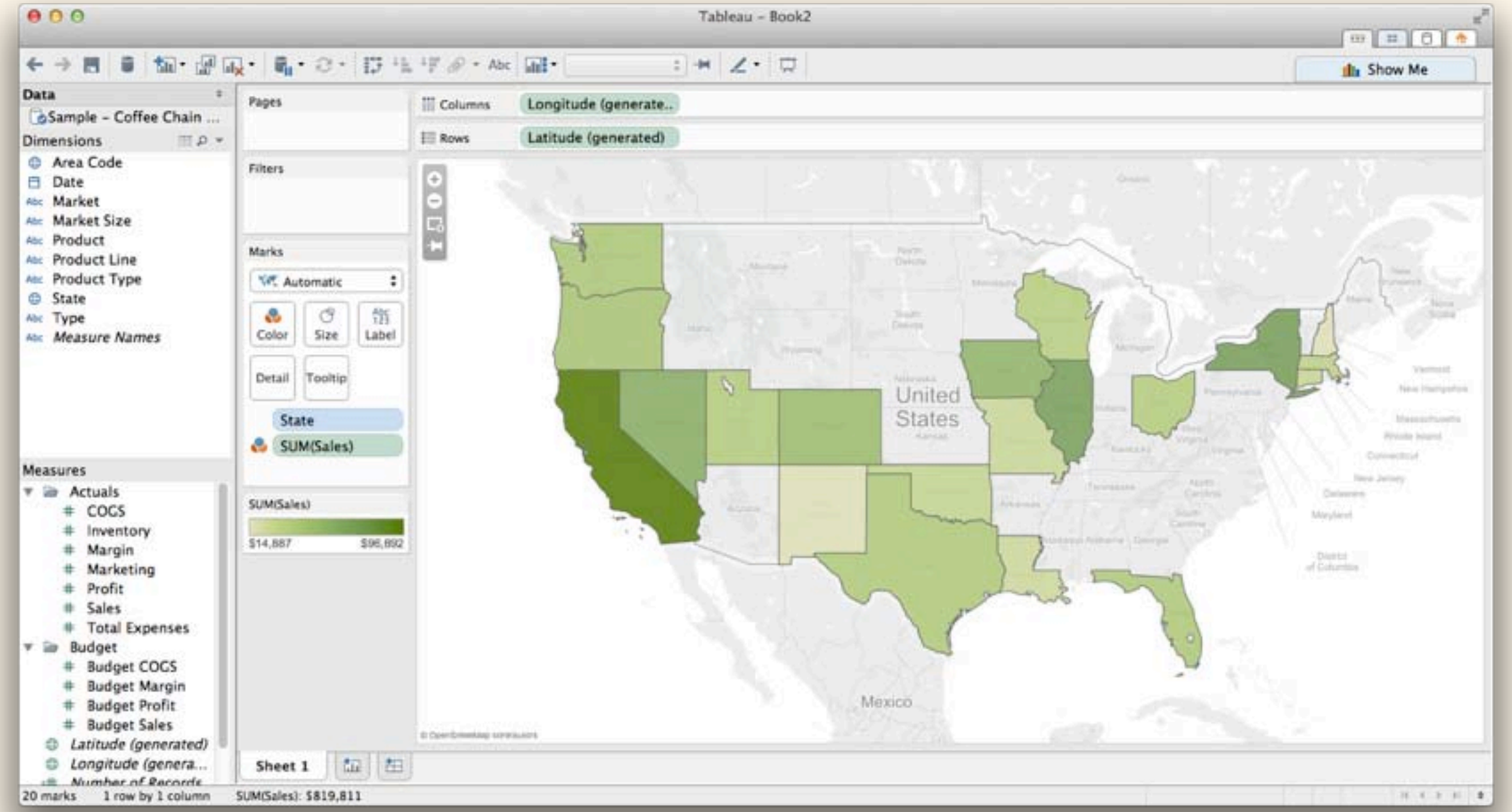
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

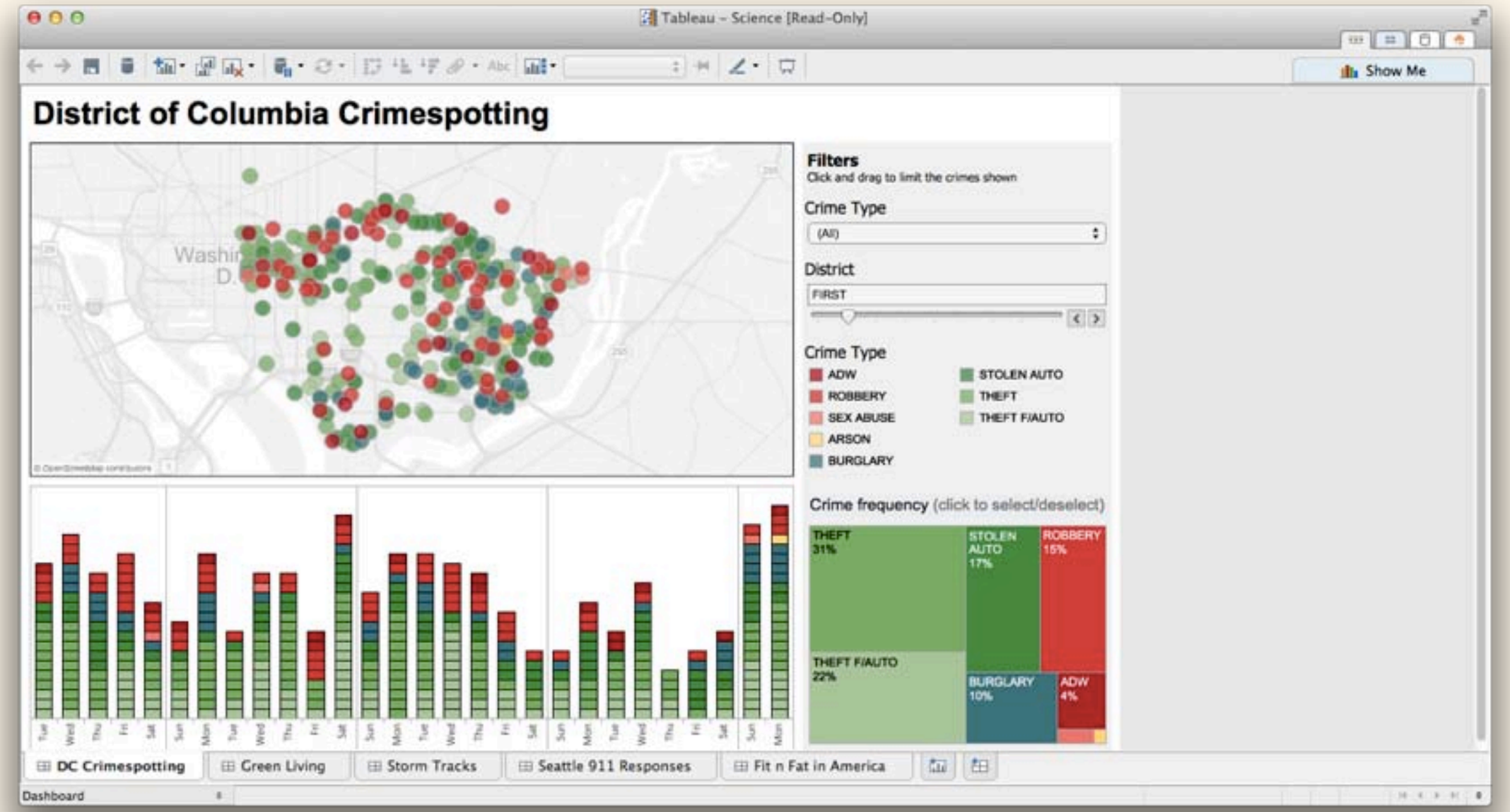
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

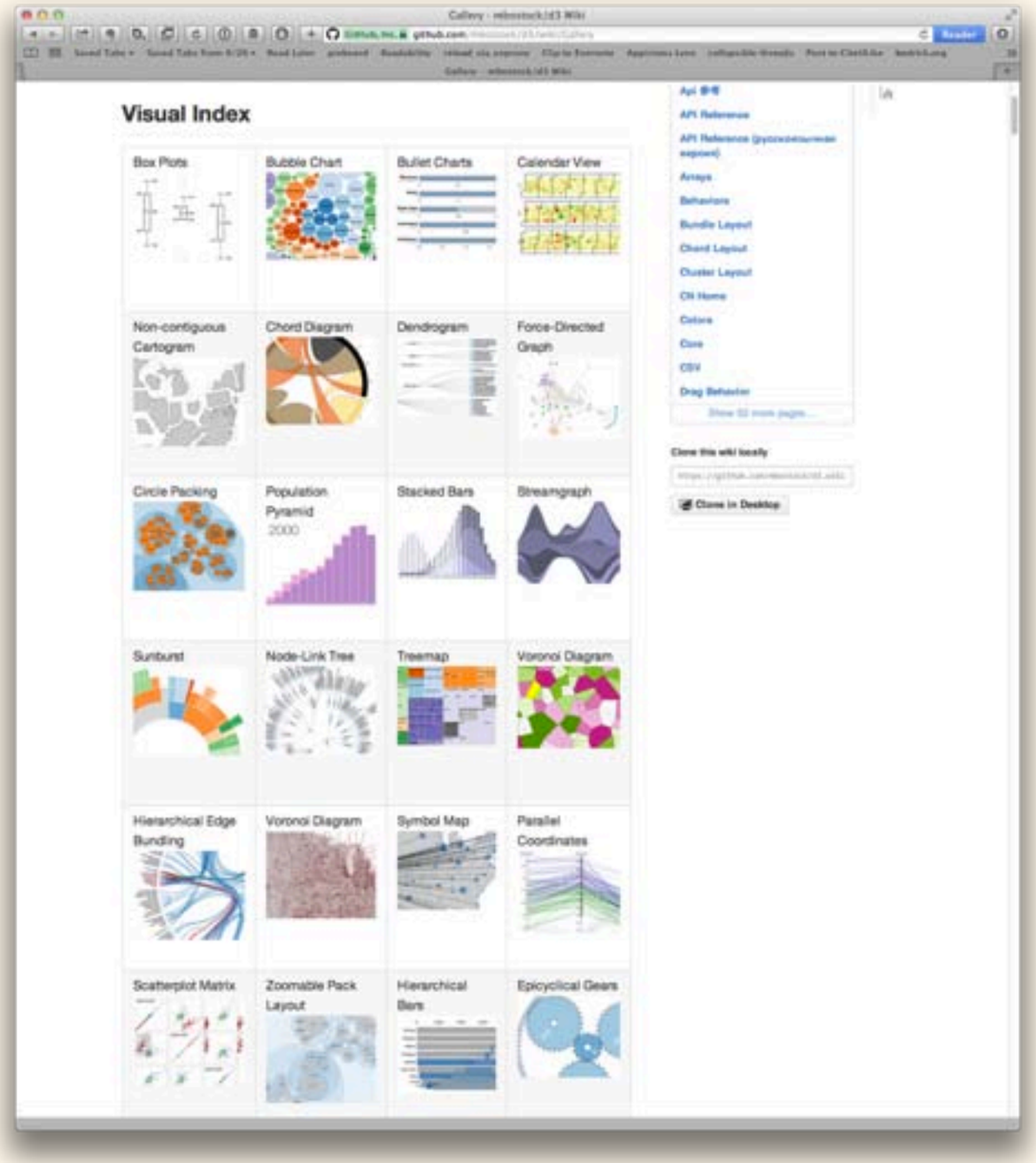
Tableau: the kitchen sink



http://www.tableausoftware.com

2. Dedicated visualization packages.

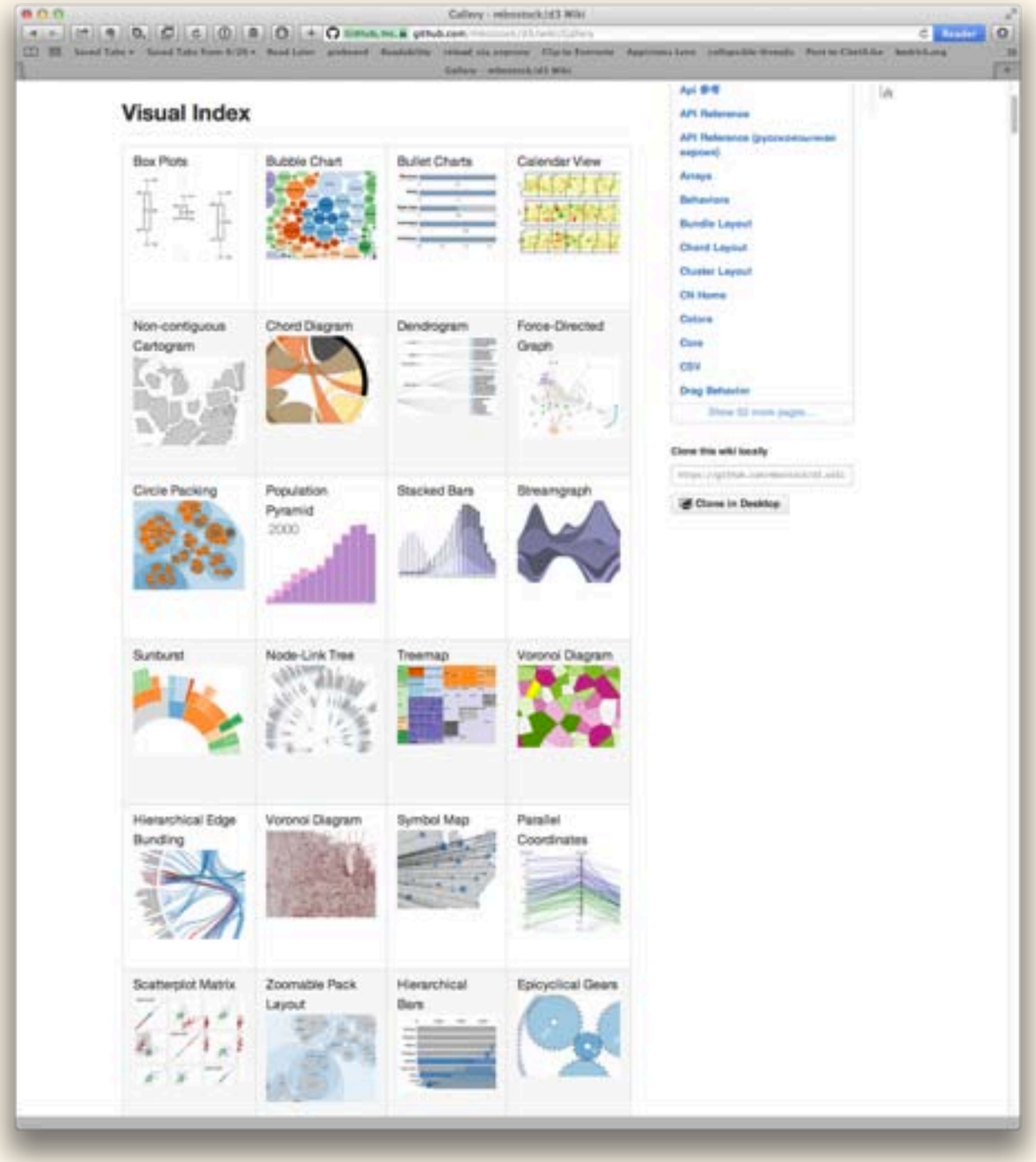
d3: Great for interactive graphics...



http://d3js.org

2. Dedicated visualization packages.

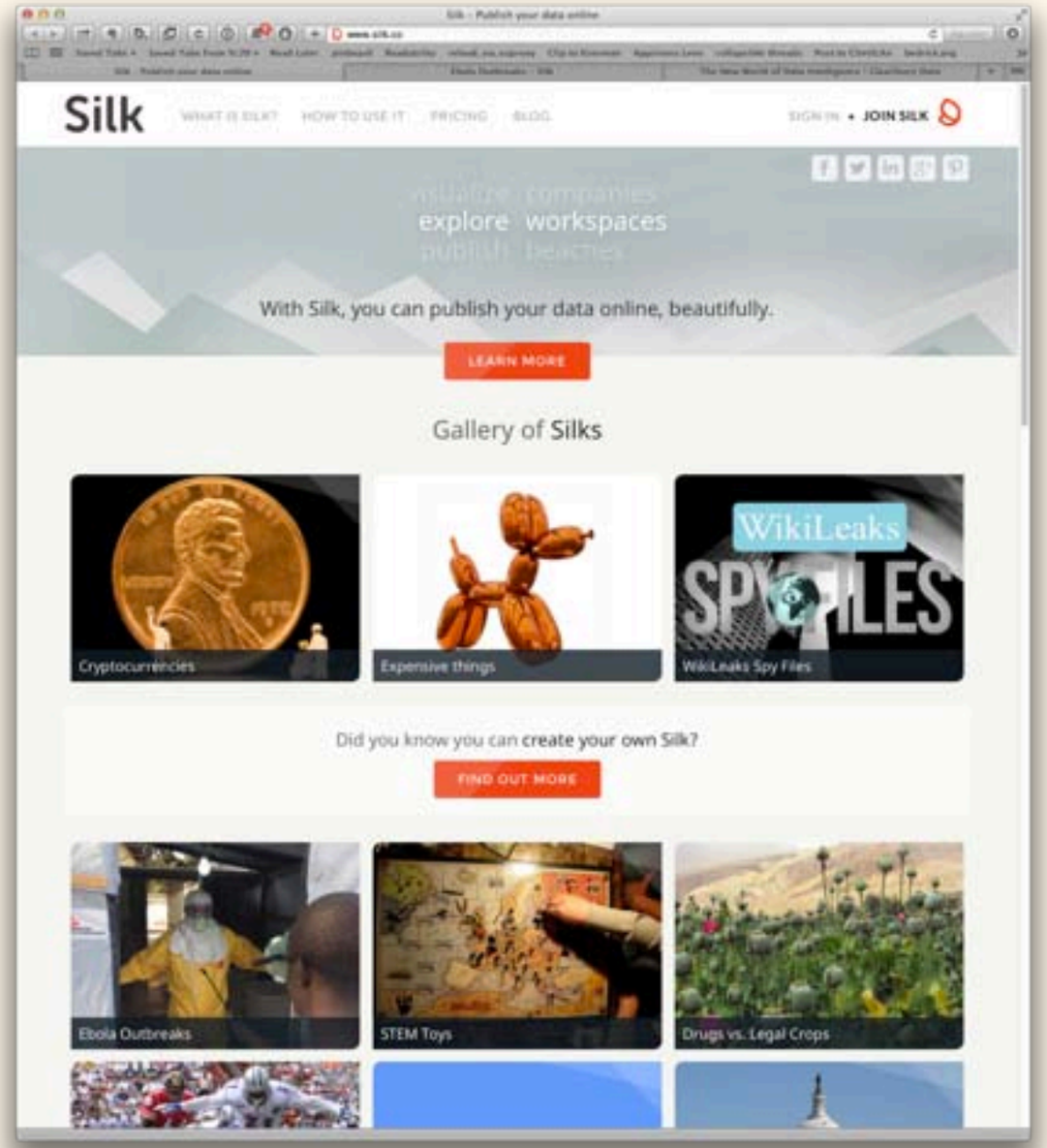
d3: ... but very programming-intensive.

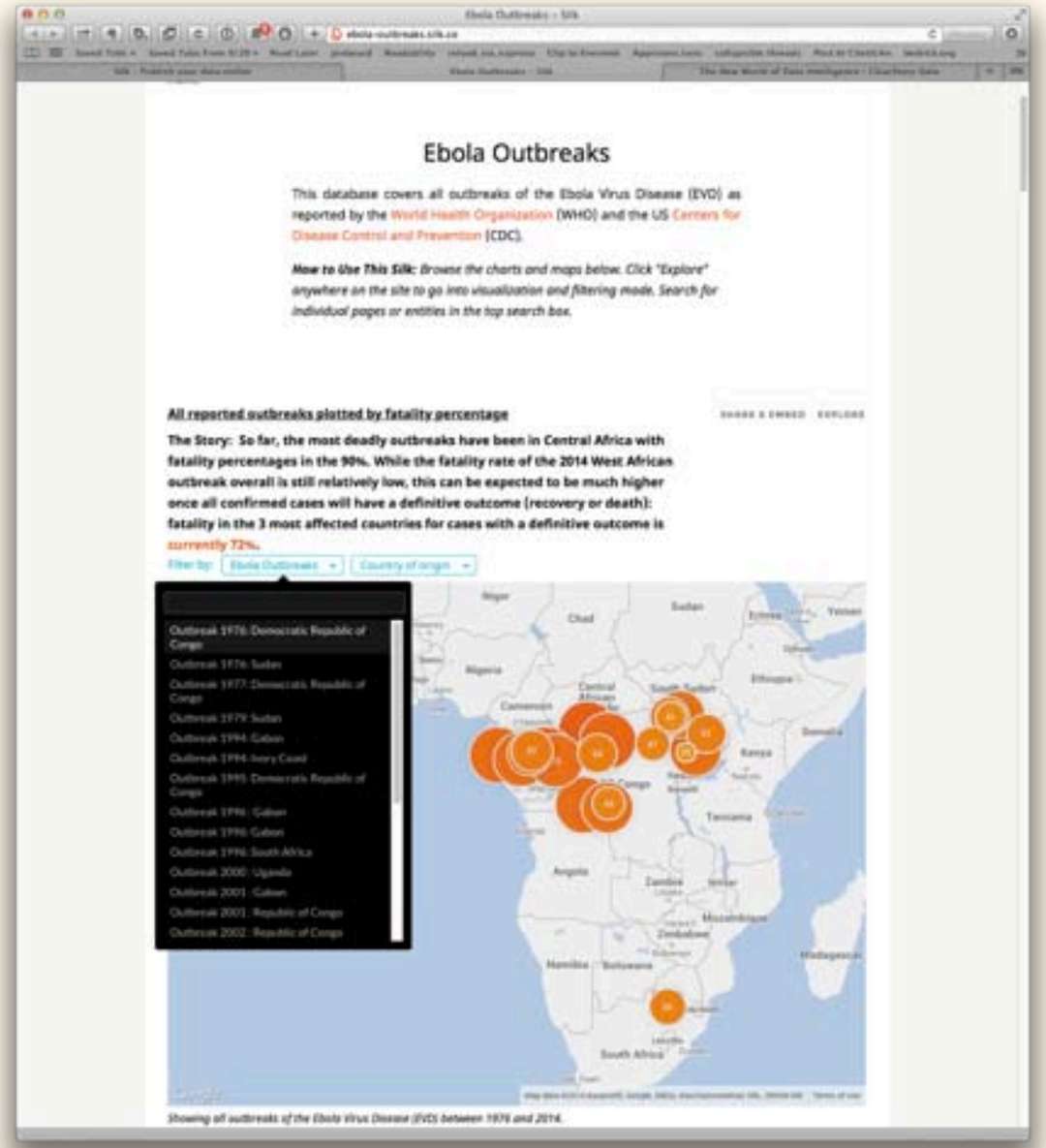


http://d3js.org

2. Dedicated visualization packages.

“Cloud” solutions (Silk, etc.):





http://www.silk.co

2. Dedicated visualization packages.

Notes:

Sci2 and Gephi are free!

Tableau is *not*!

However, it *is* free for students!

Important consideration: getting data out.

Notes on workflow:

1. Where will your graphic end up?

2. External tools

3. IP & Security

4. Change the defaults!

Notes on workflow:

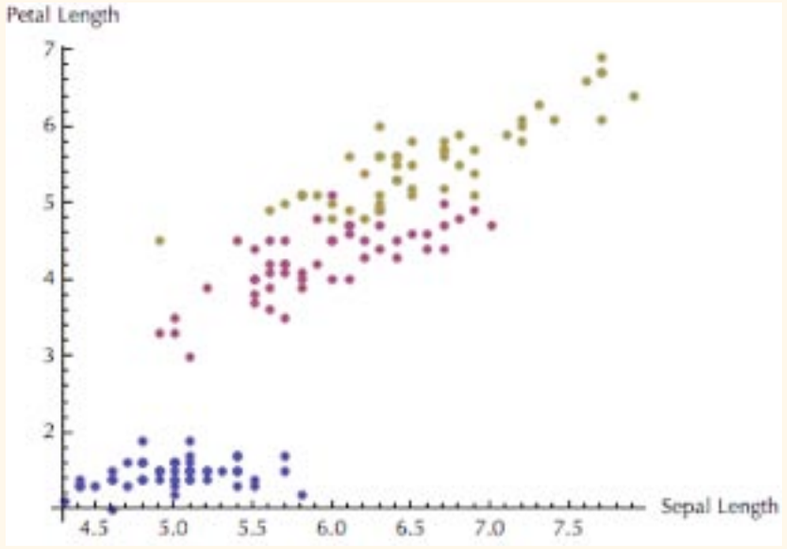
Where will your figure end up?

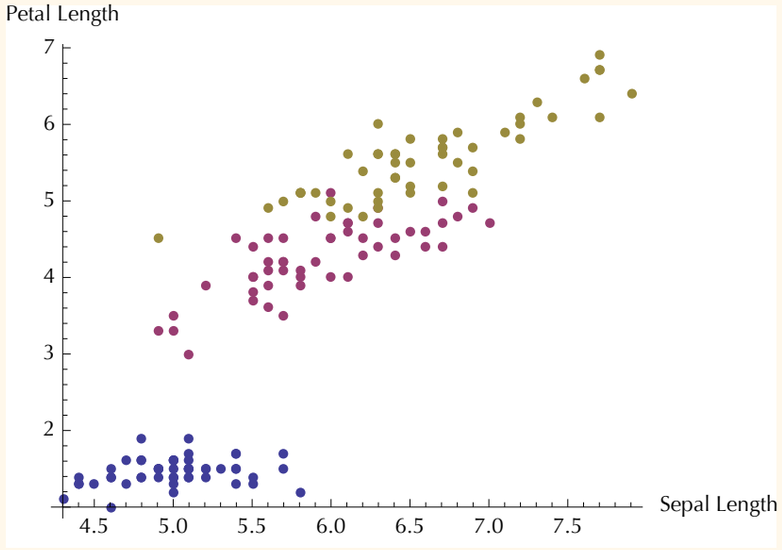
In print? In a journal? On a poster?

On a screen? Online? In a presentation?

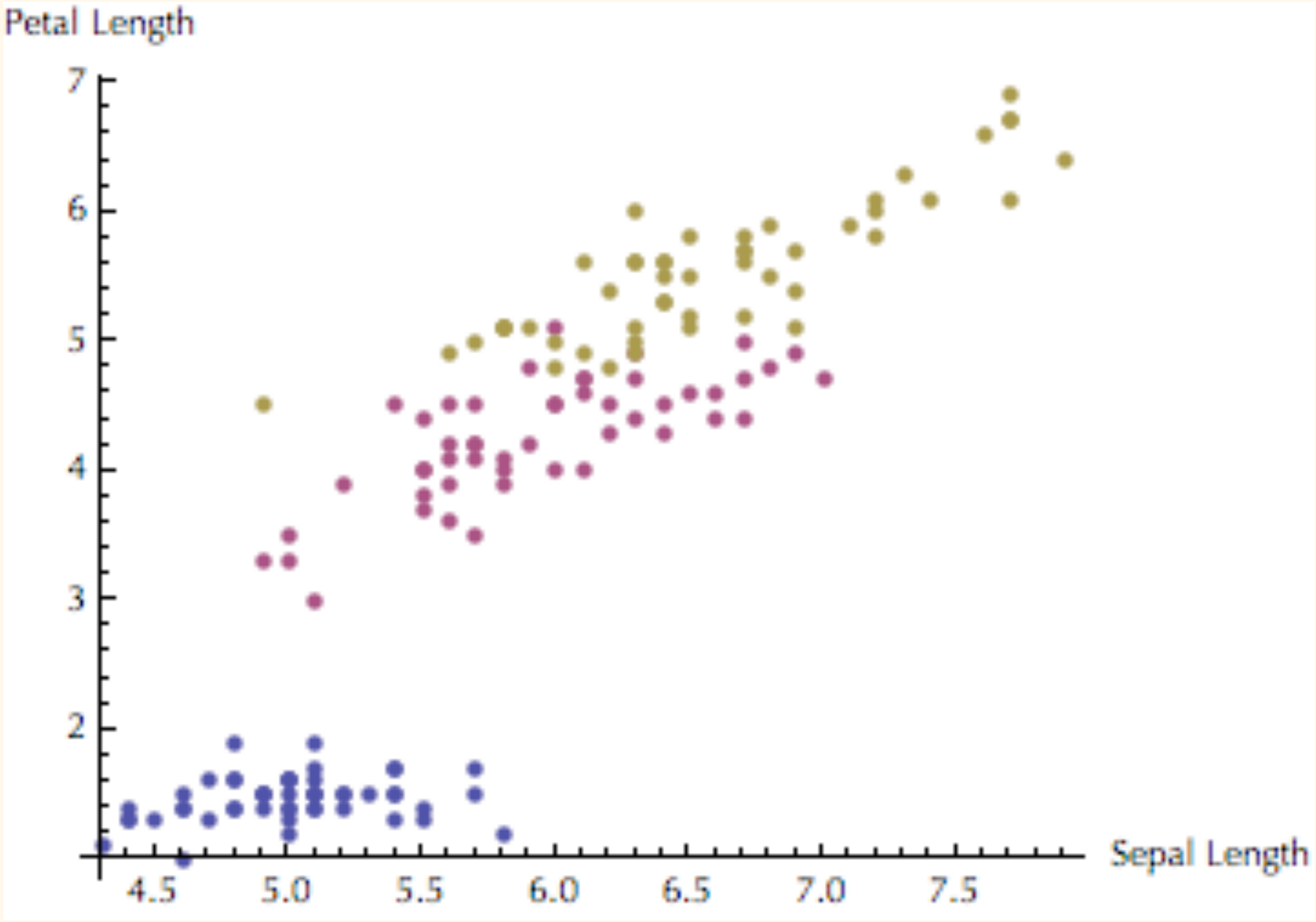
Will it be static... ... or interactive?

Vector vs. Raster images:

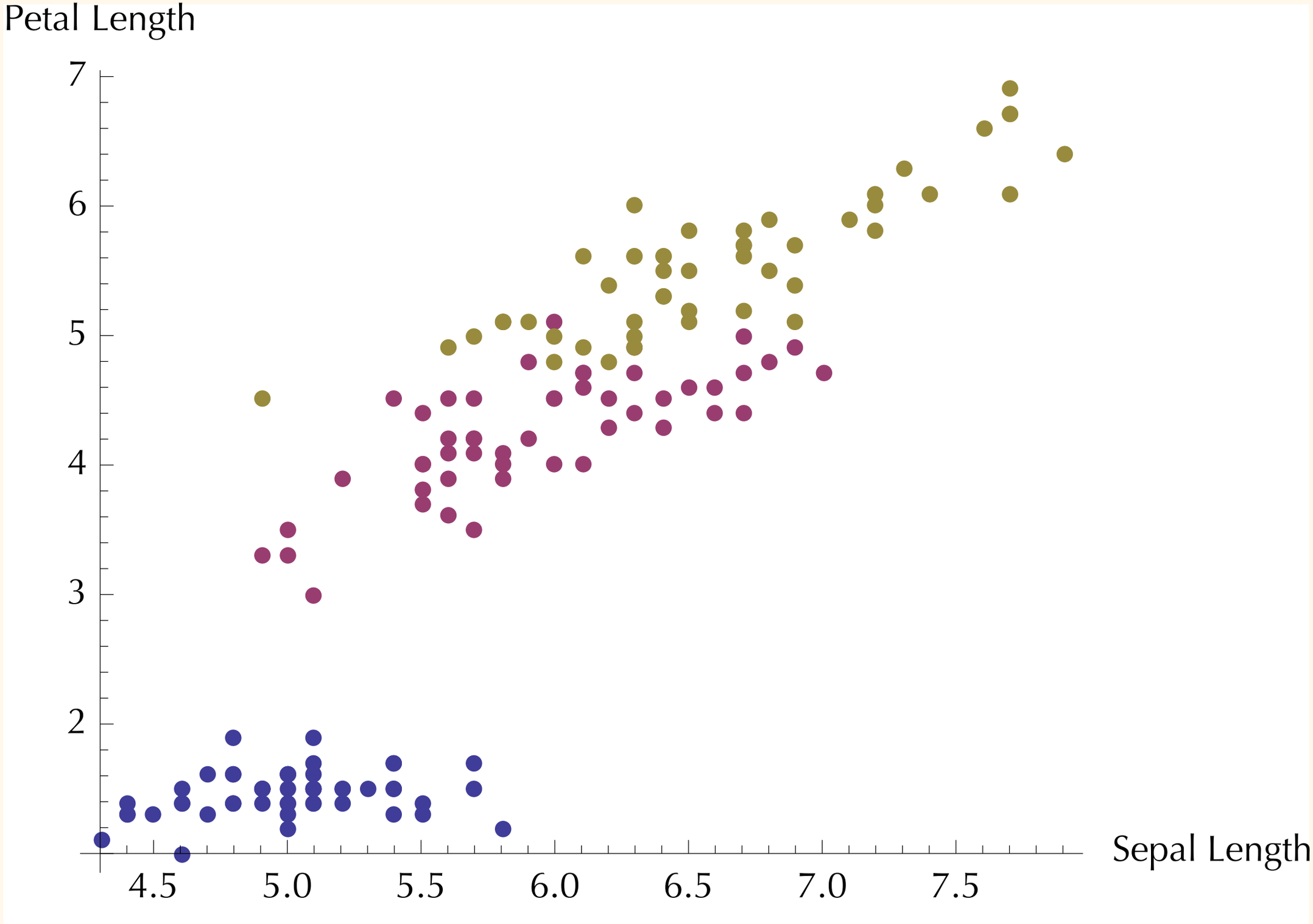




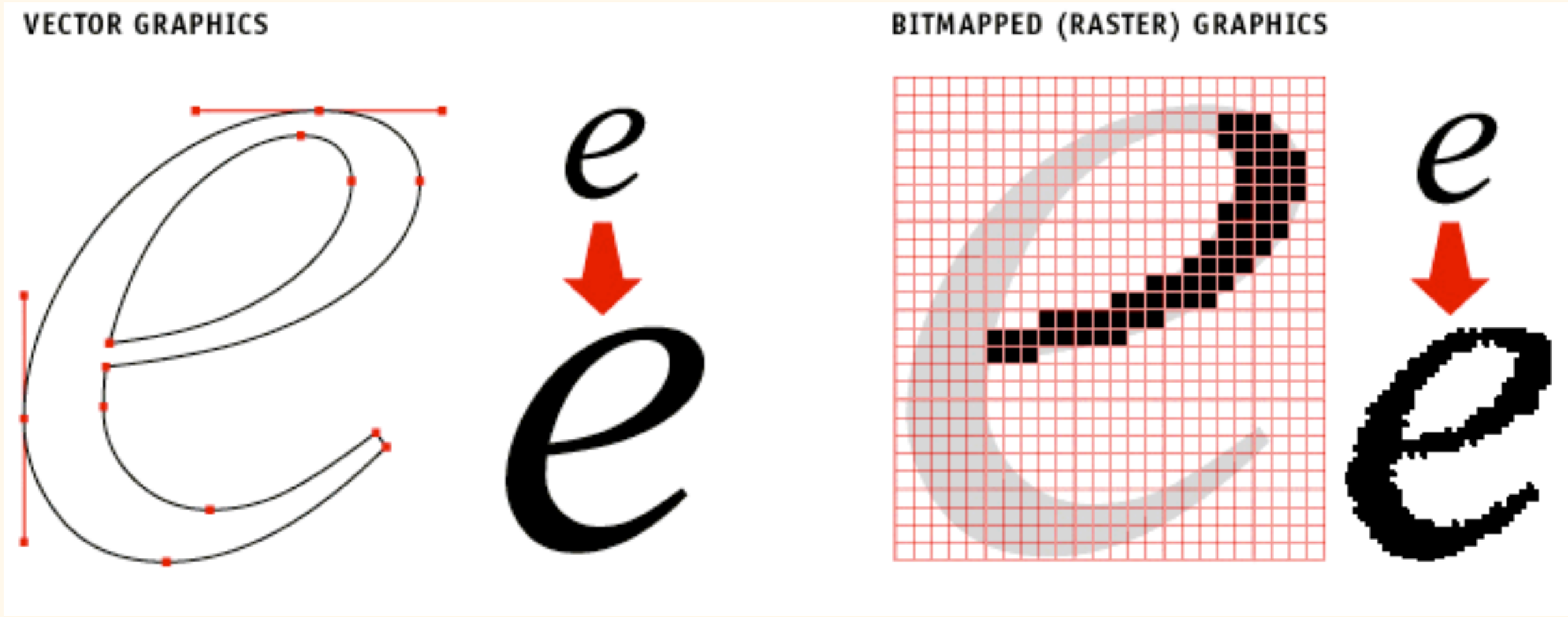
Vector vs. Raster images:



Vector vs. Raster images:



Vector vs. Raster images:



http://cornplasticstore.com/CatalogEngine/GeneralInfo.cfm?ProgramID=20522

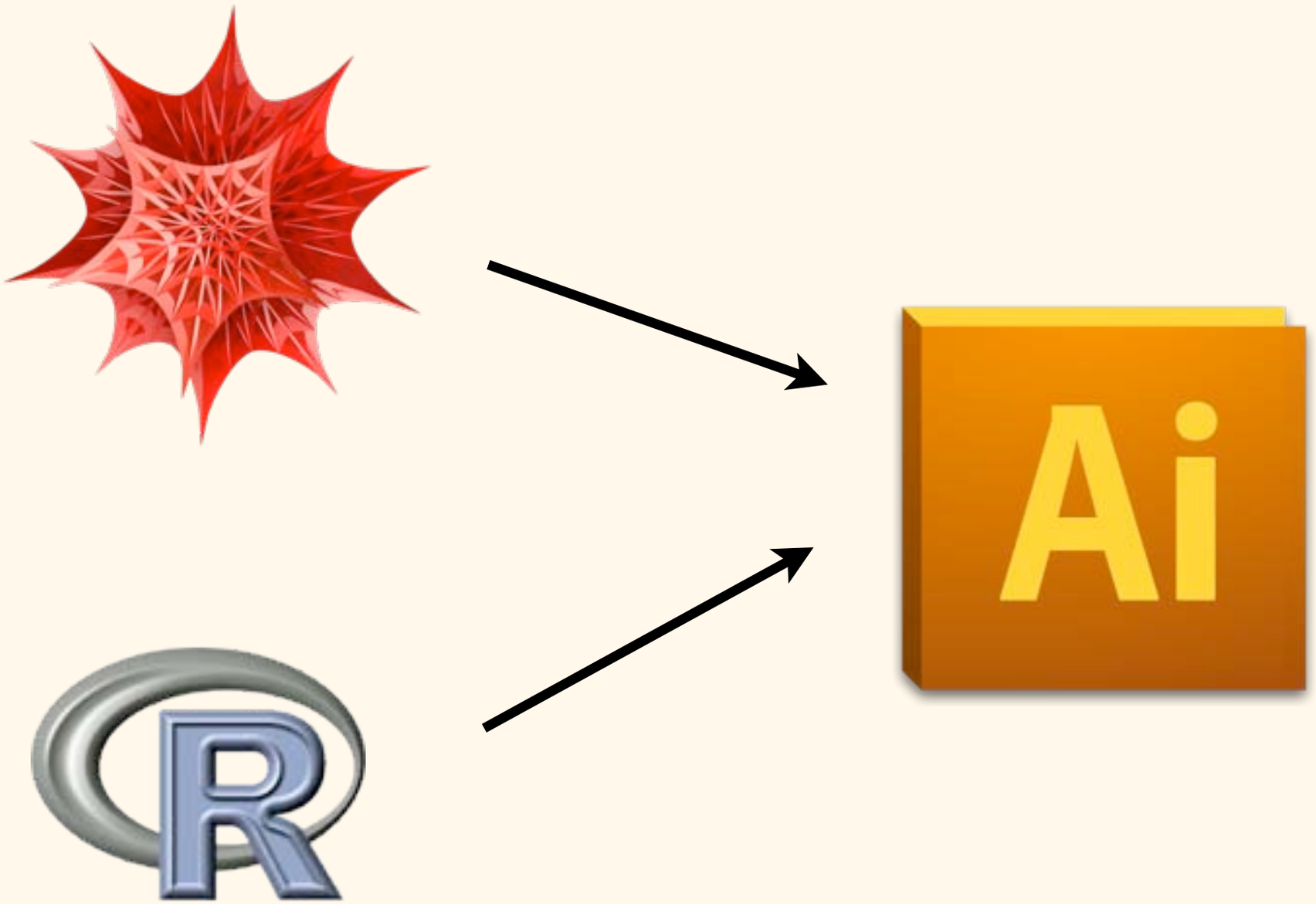
The upshot:

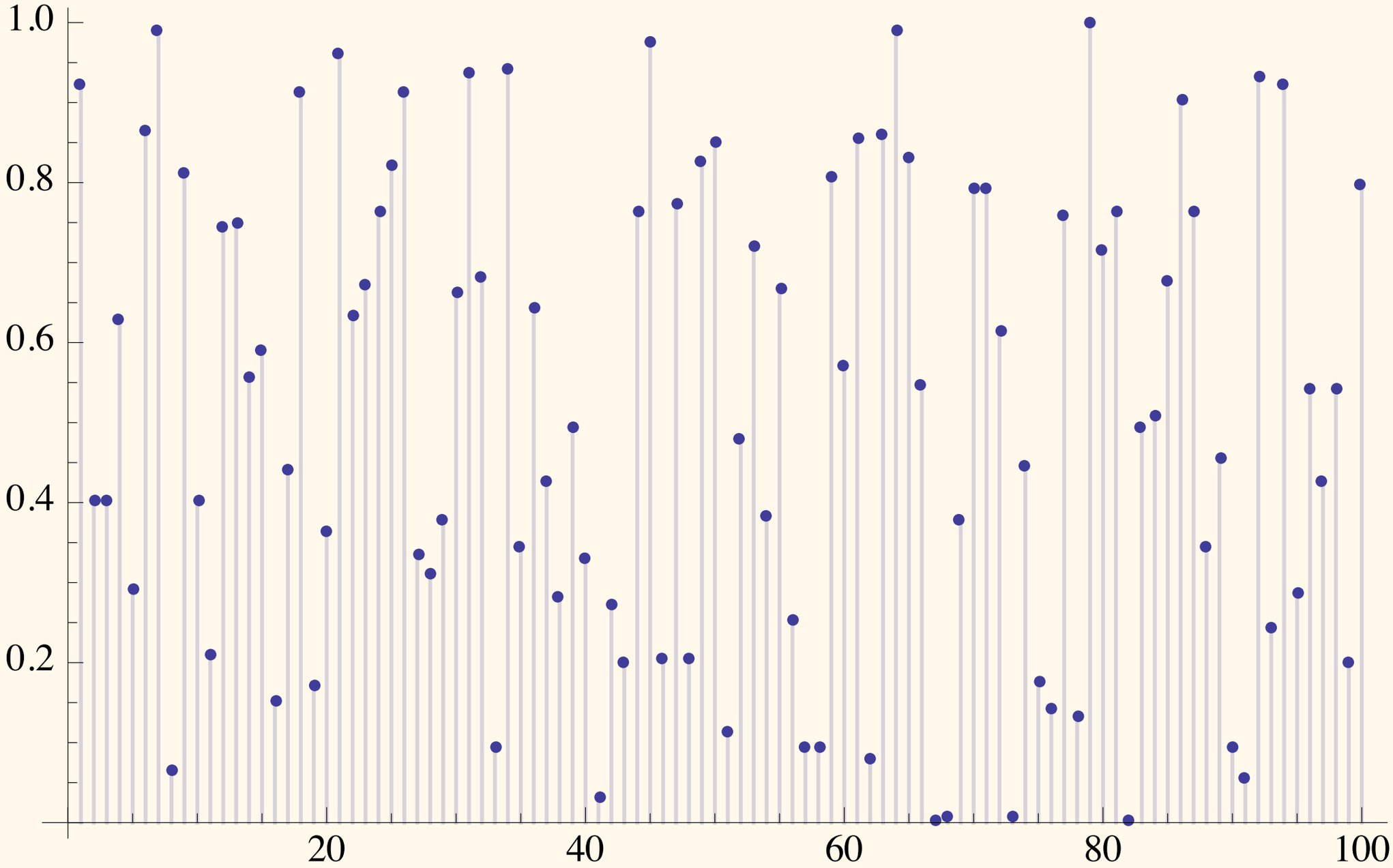
When possible, generate vector images...

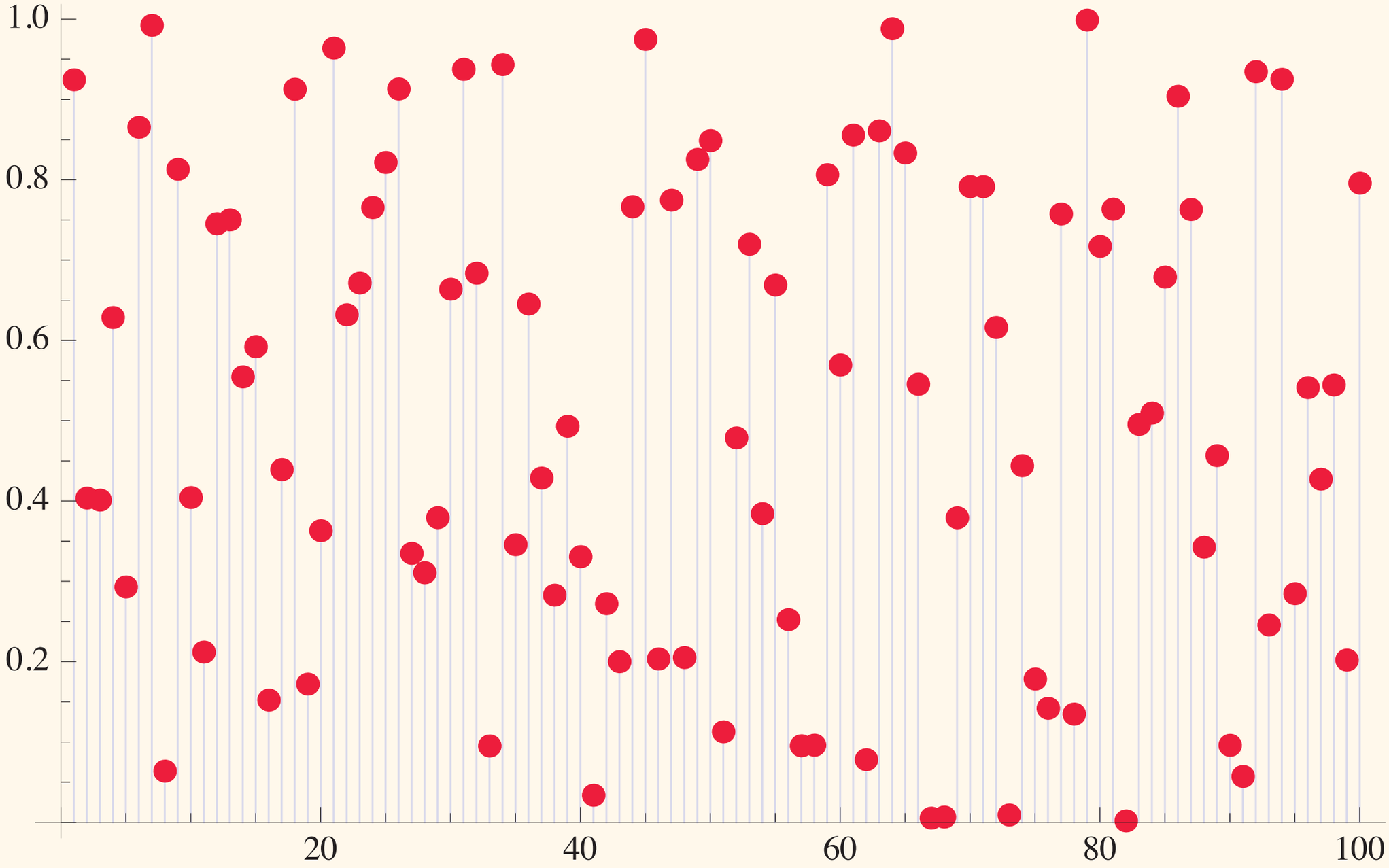
... especially for any print application!

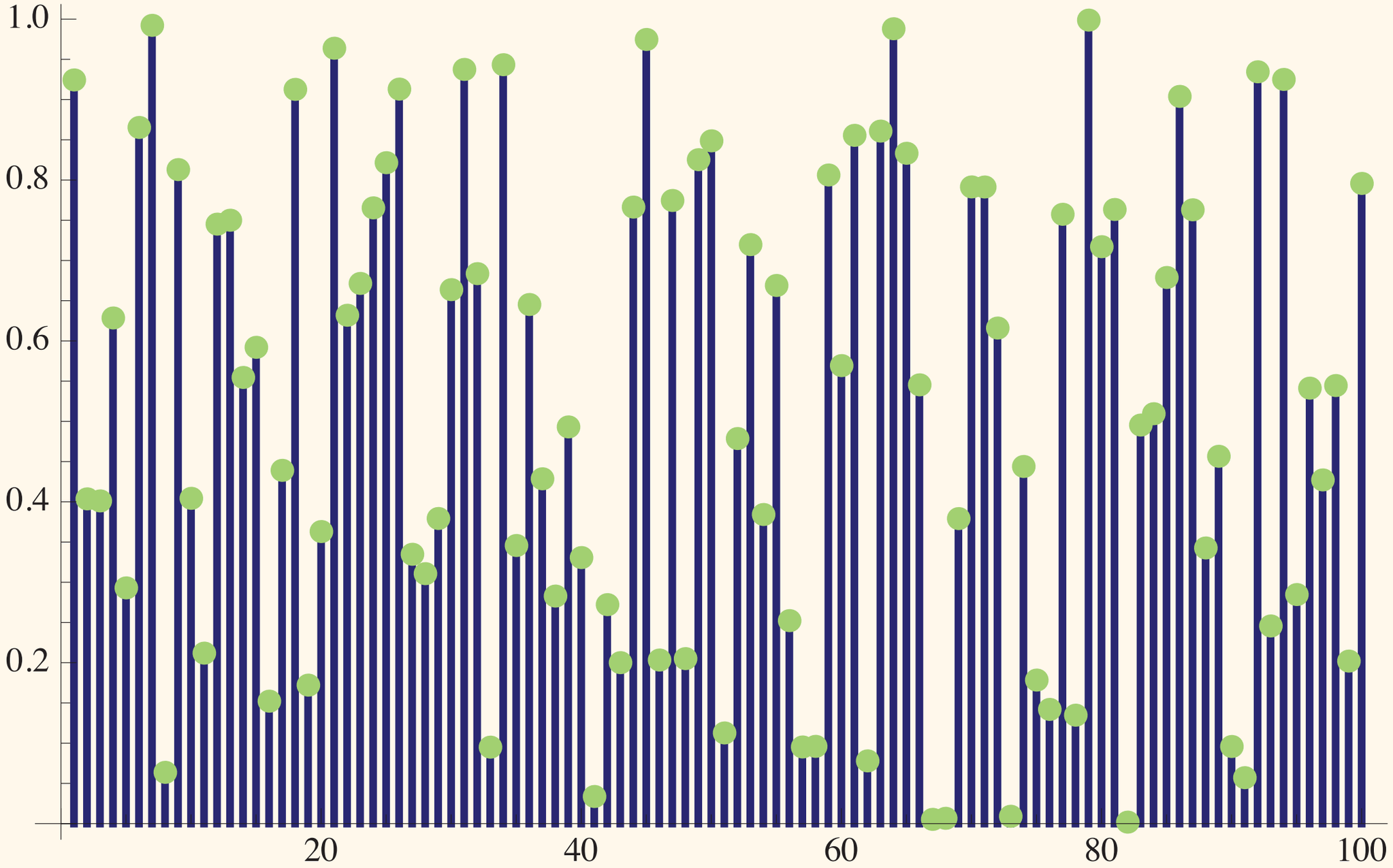
Most visualization tools (including Excel!) can do this.

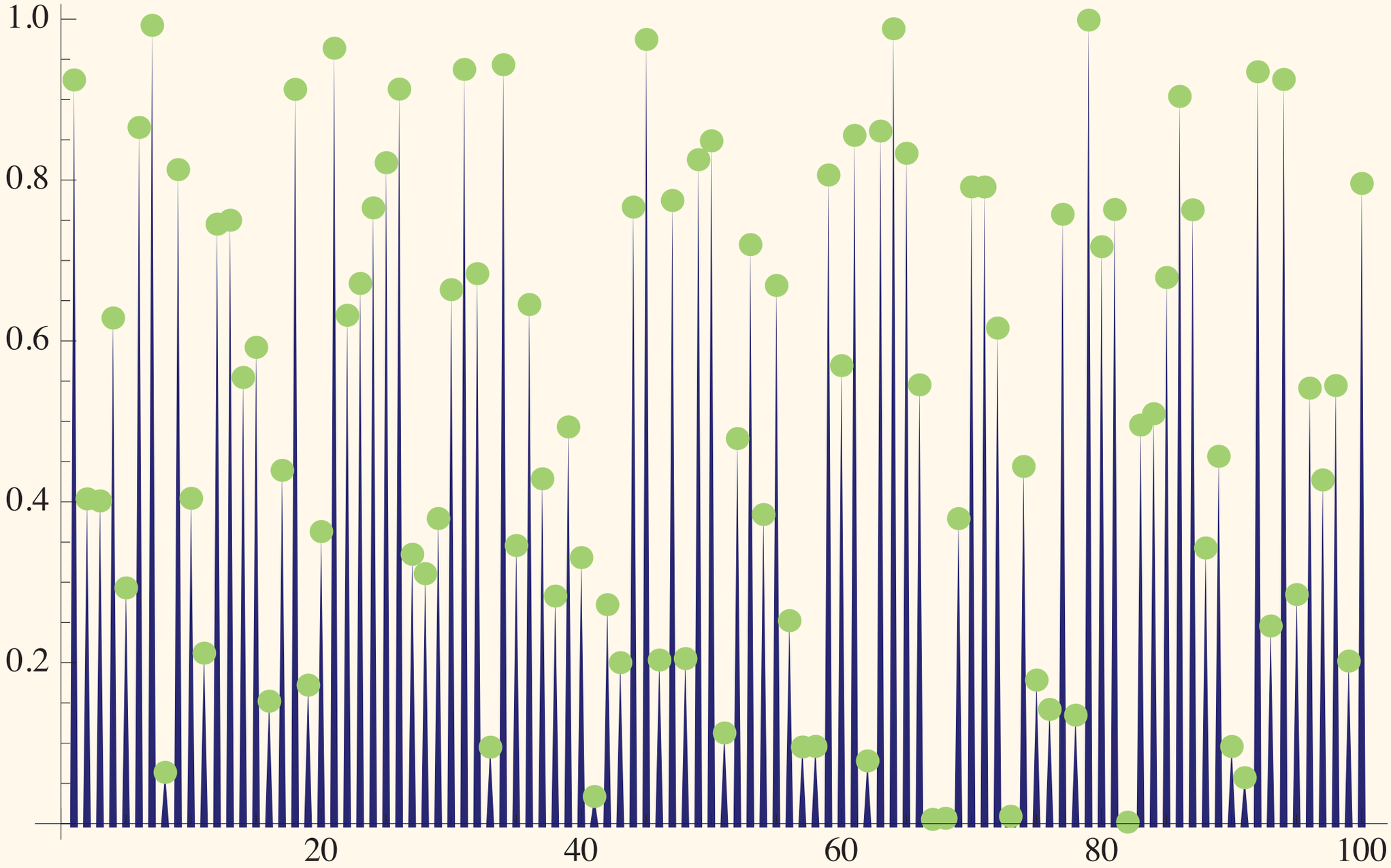
Often, multiple tools can work together.

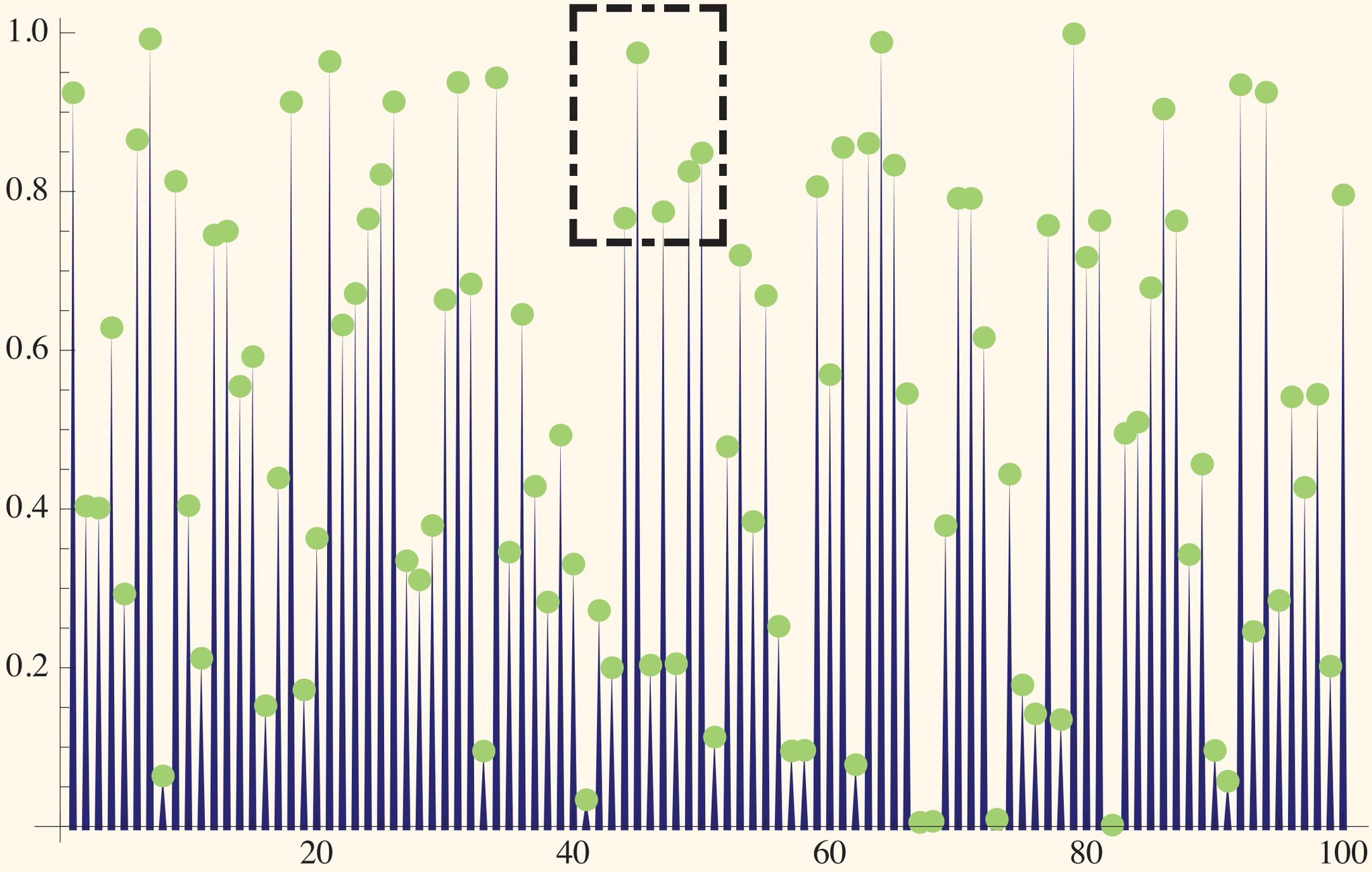



1.0
0.8
0.6
0.4
0.2
20
40
60
80
100


1.0
0.8
0.6
0.4
0.2
20
40
60
80
100

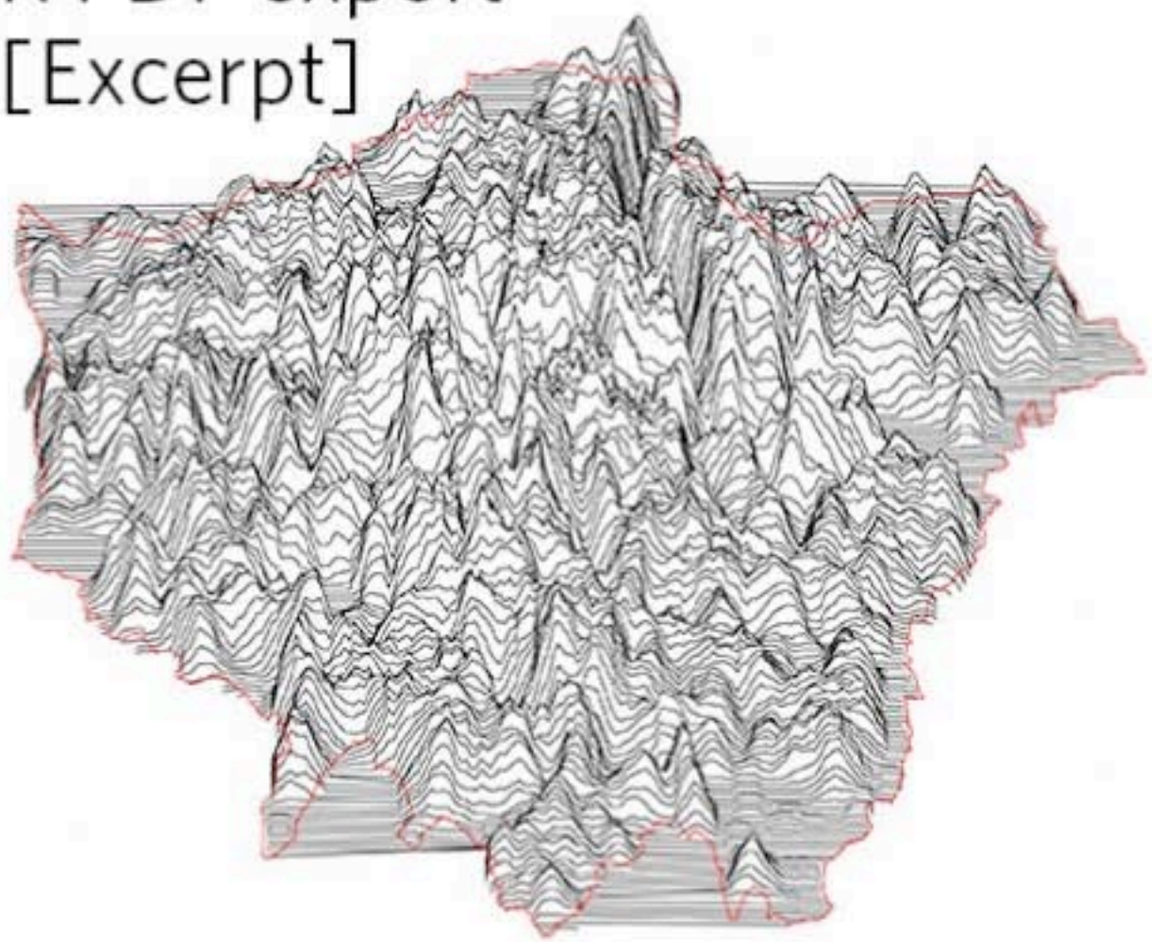

1.0
0.8
0.6
0.4
0.2
20
40
60
80
100


1.0
0.8
0.6
0.4
0.2
20
40
60
80
100

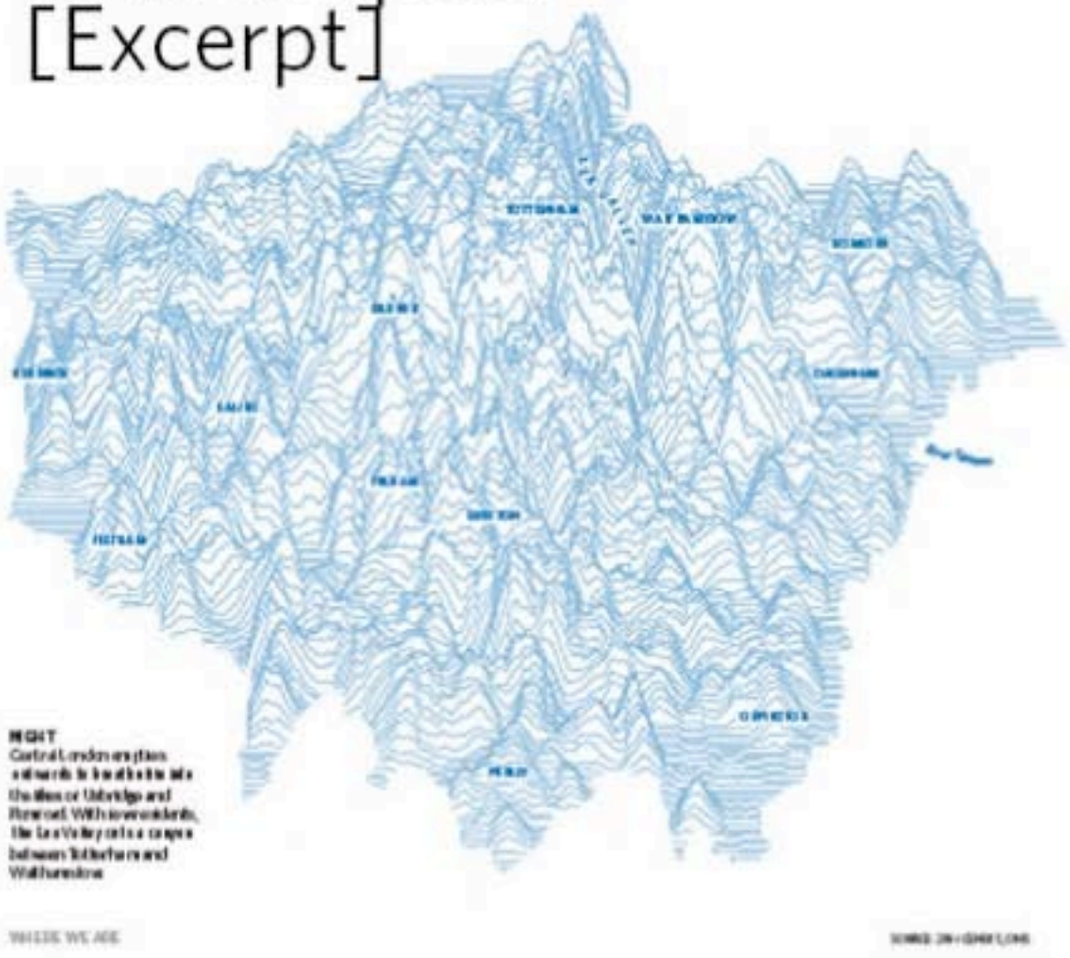

1.0
0.8
0.6
0.4
0.2
20
40
60
80
100

A less stupid example:

R PDF export
[Excerpt]



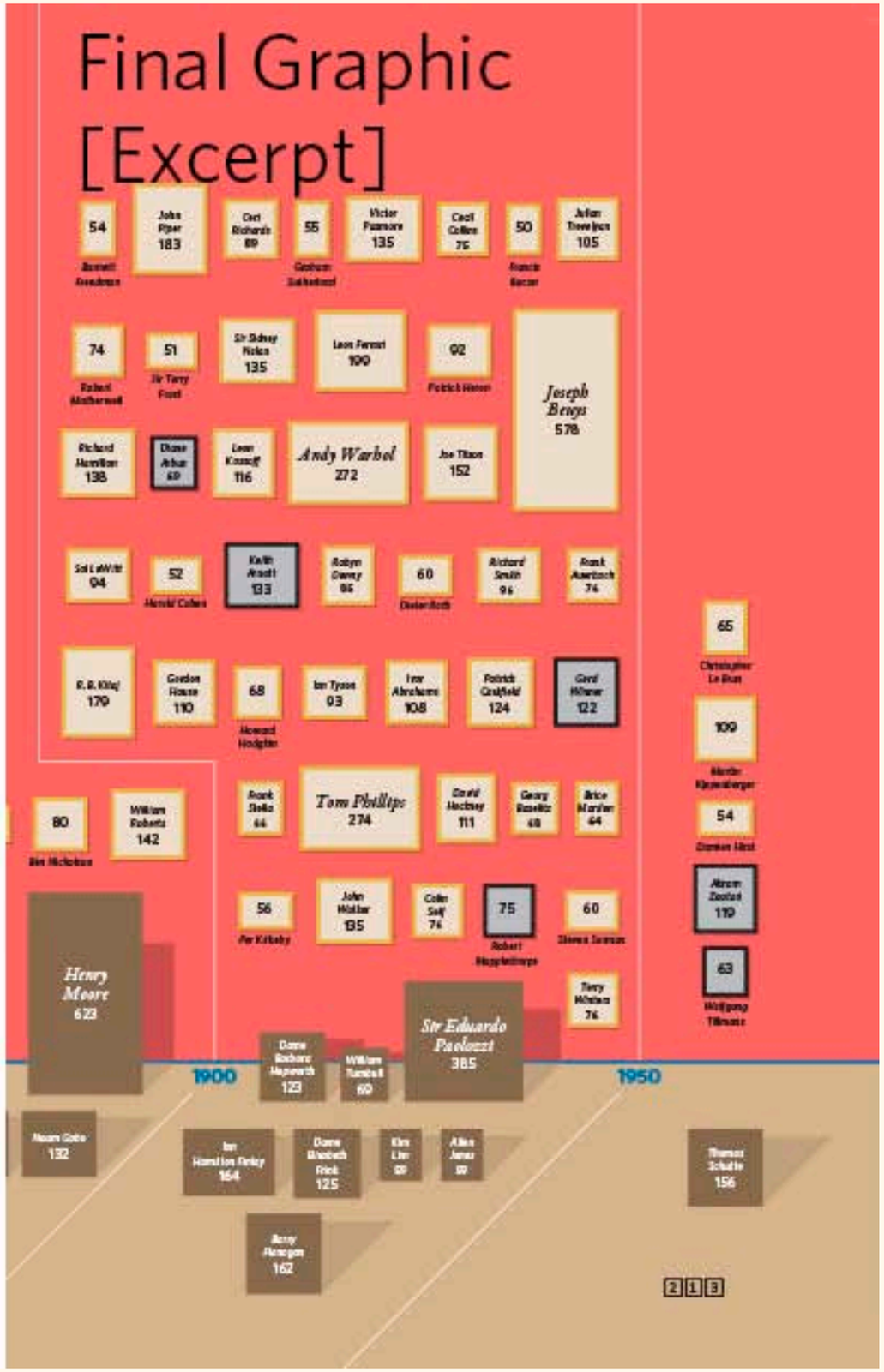
Final Graphic
[Excerpt]



Excerpted from *London: The Information Capital* by James Cheshire and Oliver Uberti (Particular Books, 30 October 2014)

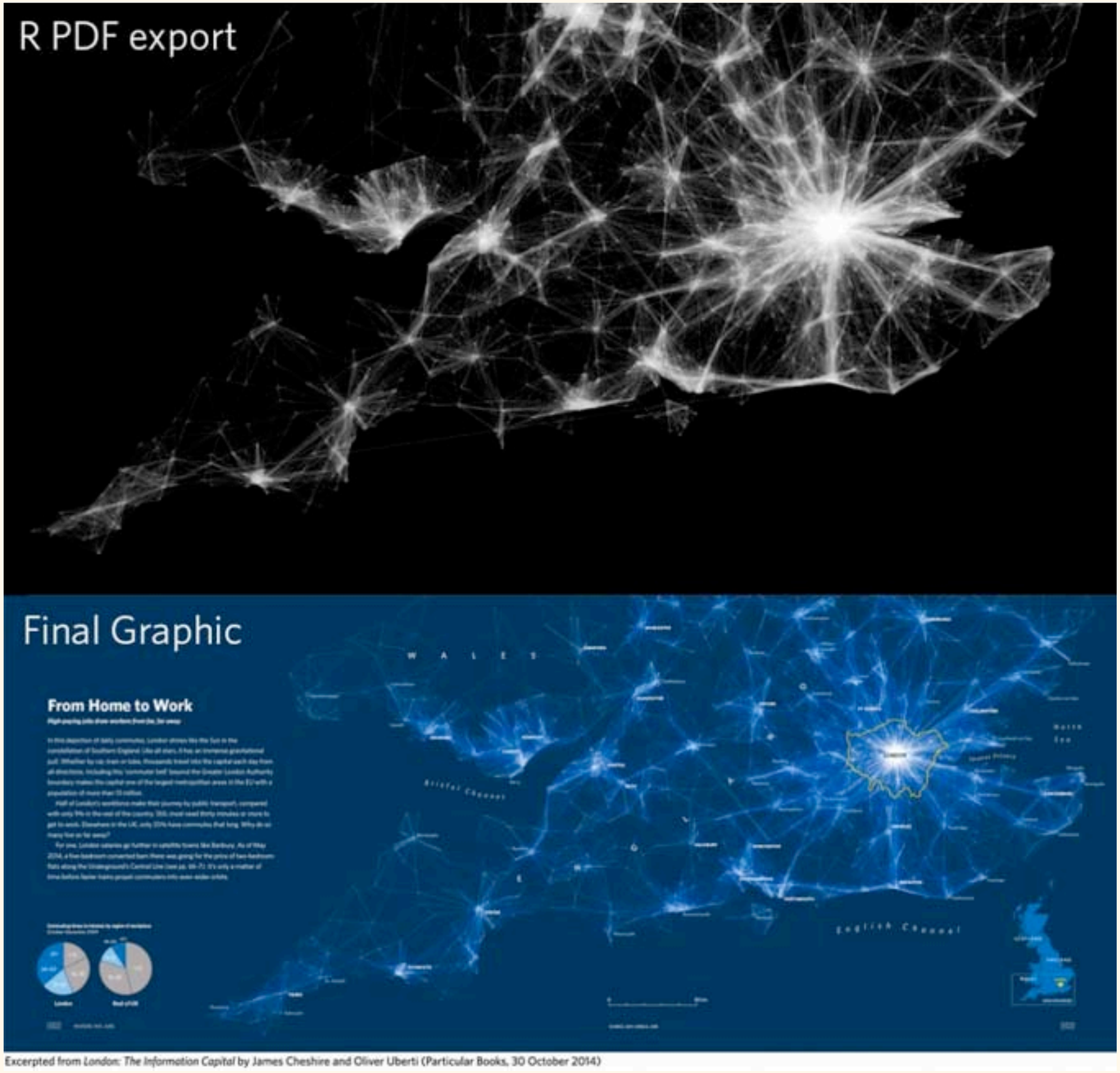
http://theinformationcapital.com/coder-designer/

A less stupid example:



http://theinformationcapital.com/coder-designer/

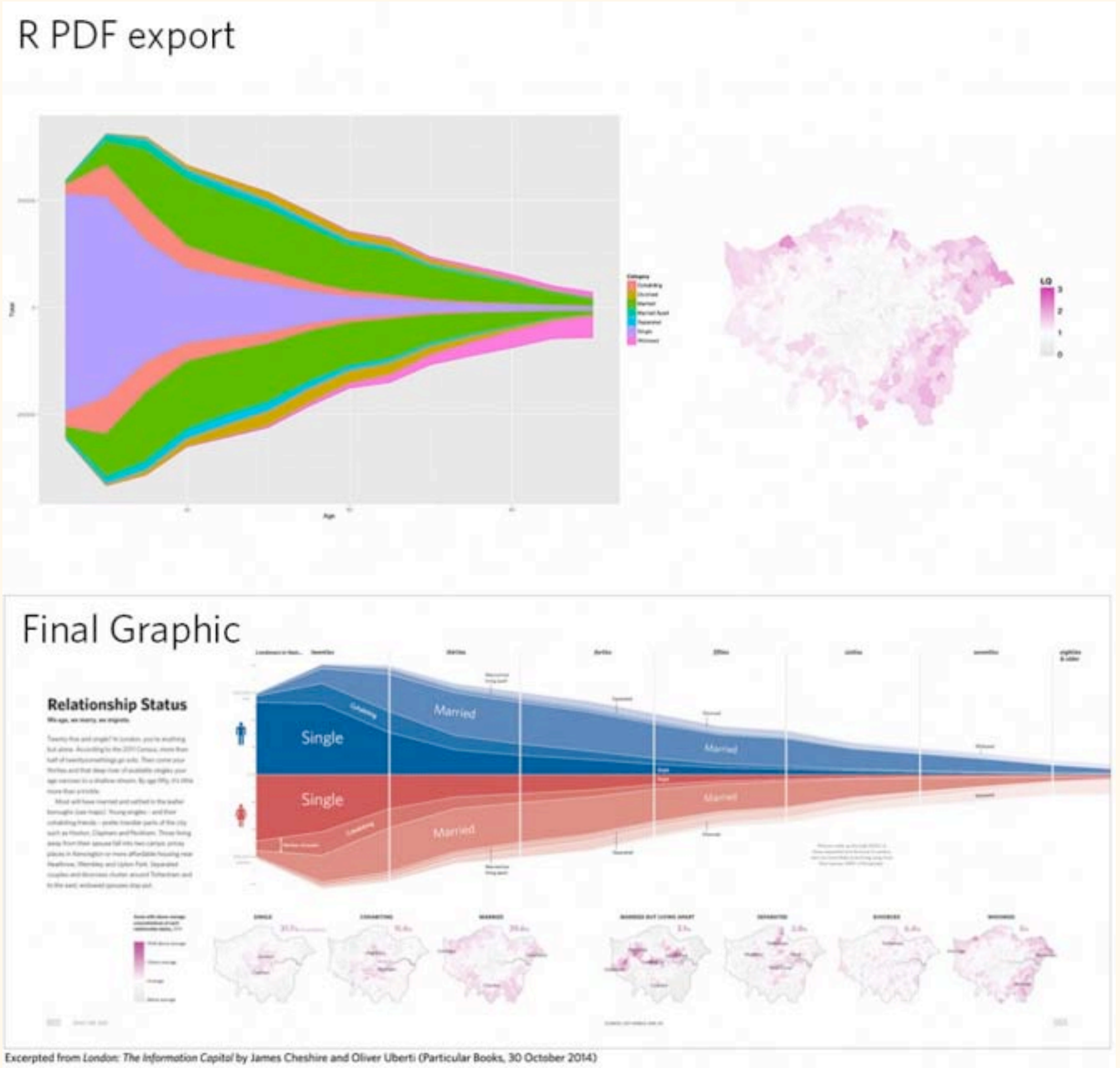
A less stupid example:



Excerpted from *London: The Information Capital* by James Cheshire and Oliver Uberti (Particular Books, 30 October 2014)

http://theinformationcapital.com/coder-designer/

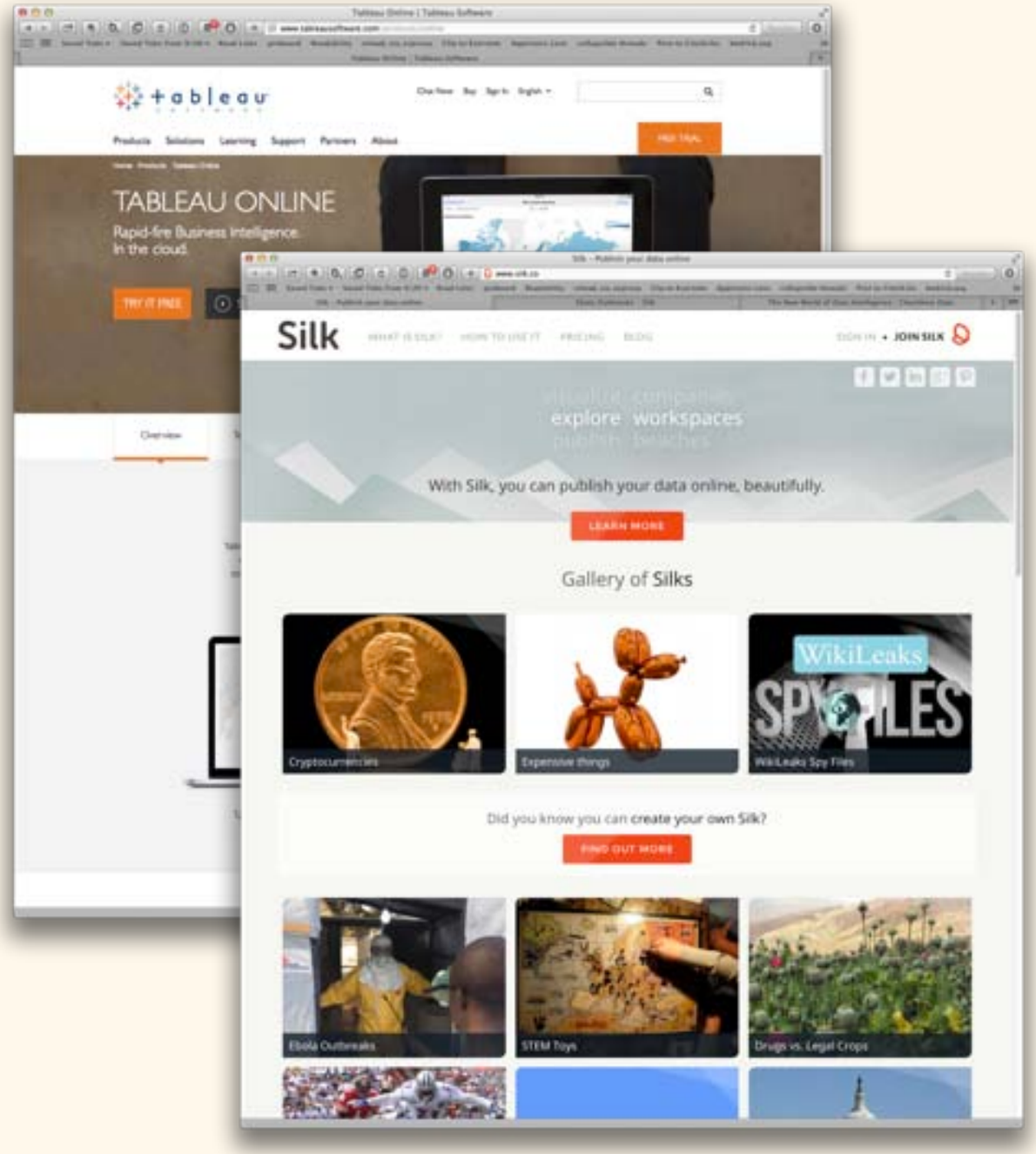
A less stupid example:



Excerpted from *London: The Information Capitol* by James Cheshire and Oliver Uberti (Particular Books, 30 October 2014)

http://theinformationcapital.com/coder-designer/

IP & Security:



Think before you upload!

Terms of use: “What am I giving away?”

Unpublished data?

IRB & HIPAA?

When in doubt, ask your neighborhood librarian!

Change the defaults:

Your tool's default settings are *not* the final word.

Small changes can have a large cumulative effect!

Example: Jackie's competitive eating chart!