Building a

Stamman for

statistical graphics

in Clojure

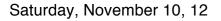
Kevin Lynagh Keming Labs @lynaghk

2012 November 9 Øredev Malmö, Sweden

Agenda



Data Visualization





Grammar Graphics



Data Visualization what + why + how

Wind energy

Turbine: 18 ven/winergy Speed: 1463.74 RPM Power: 1450.00 kW								
Main Bearing Venation (ing	5	100	alo	obe	ebo	sóo	600	
Ring Gear Vibration (mg)		100	200	300	400	sóo	600	
Internediate Speed Shaft Generator End Bearing Vibration (mg	P	100	200	obe	400	sóo	600	
High Speed Shaft Rotor End Bearing Vibration (mg	Į.	100	200	sio	400	500	600	
High Speed Shaft Generator En Bearing Vibration (mg	ĺ.	100	200	300	450	600	600	
Generator Drive End Bearing Vibration (mg	-	160	260	300	400	600	600	
Generator Non-Drive End Bearly Vibration (ing)	, F	160	alo	ode	eòo	sdo	600	
OI Particle count	Į,	łò	16	-	1000	10000	100000	

Bioinformatics / EdgeBio

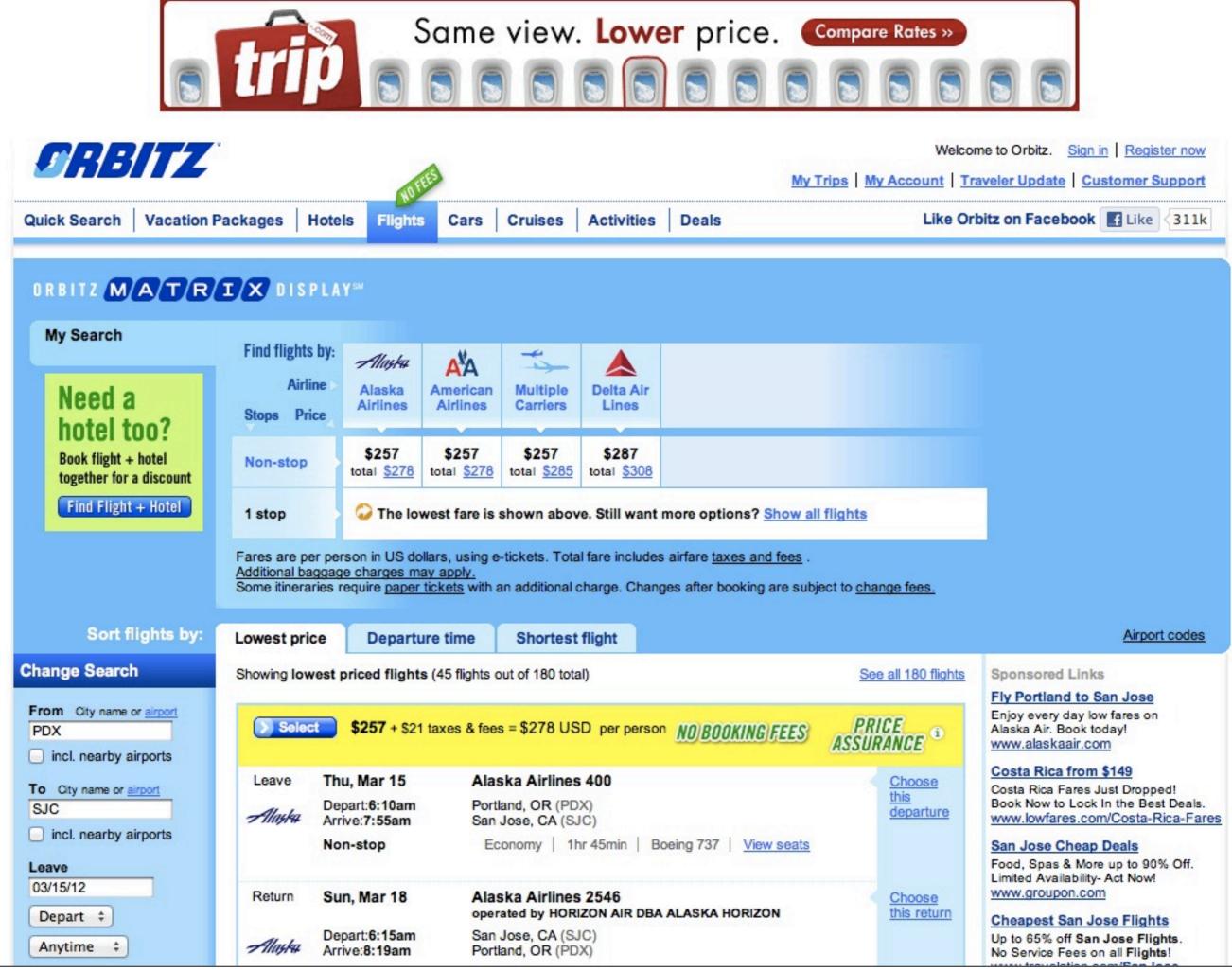
🗯 Chrome File Edit View History Bookmarks Window Help	🗈 🖯 👫 5 🎲 🌲 🗖 🕙 🕴 📿 🖣 🗹 🖬 🖬 (1:22) Sat 2:31 PM Q
● ○ ○ / — localhost:8080 ×	
← → C 🗋 localhost:8080	☆ 🔣 🖳 💿 🥥 🍕
VCF Visualizer	Logged in as: kerninglabs
VCF files	
Select VCF files	
Export subset	

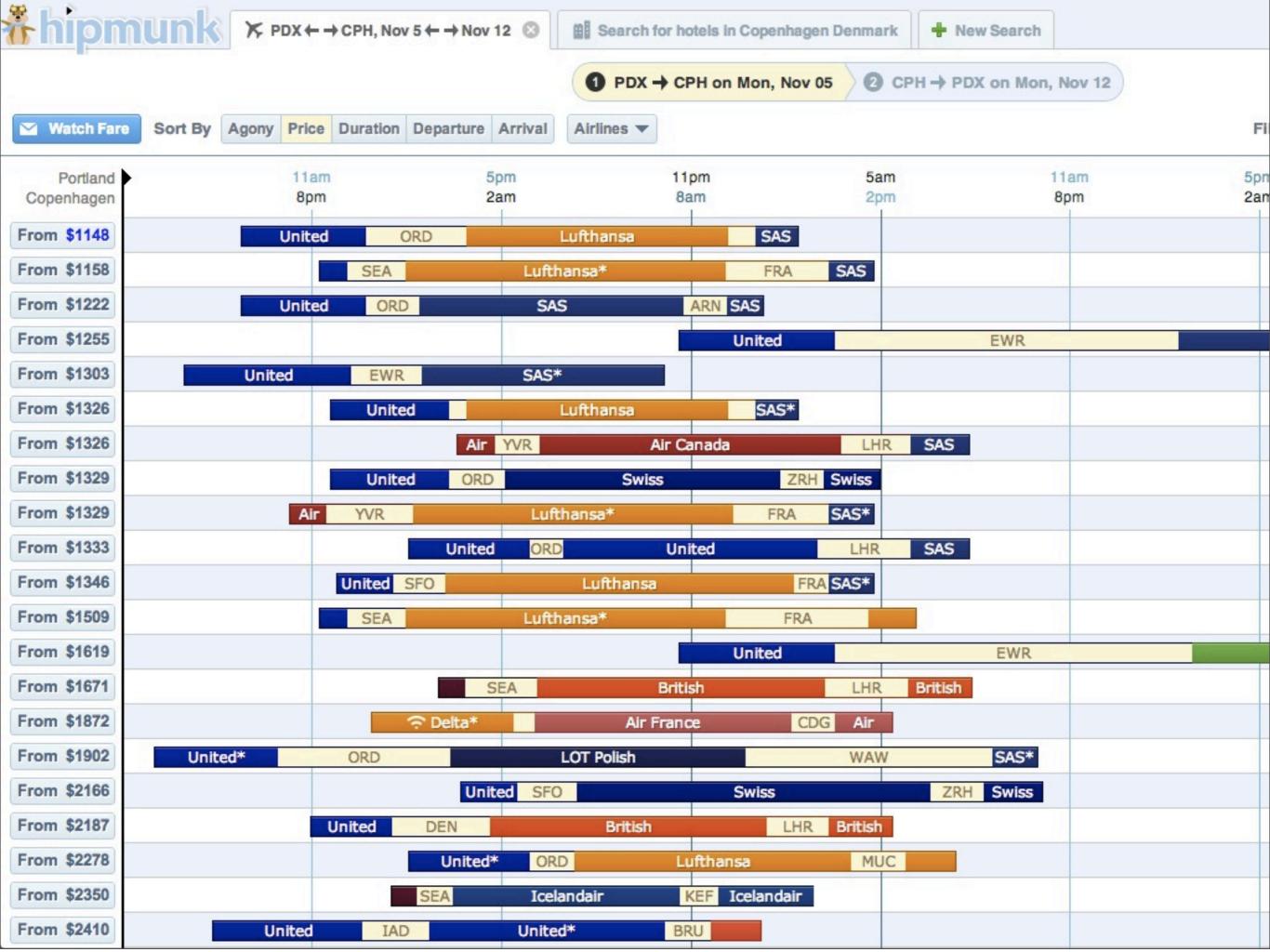
VE RI ES

Doc & patient,

meet



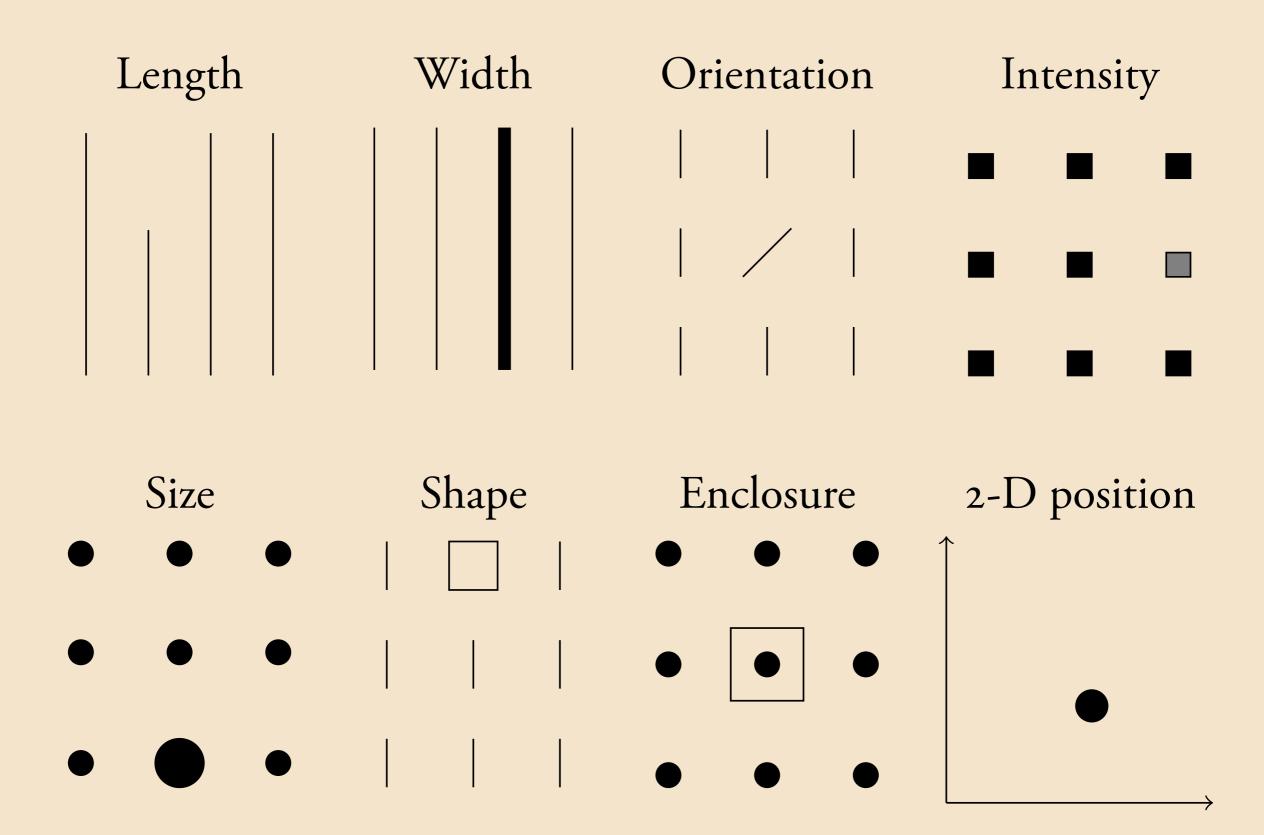


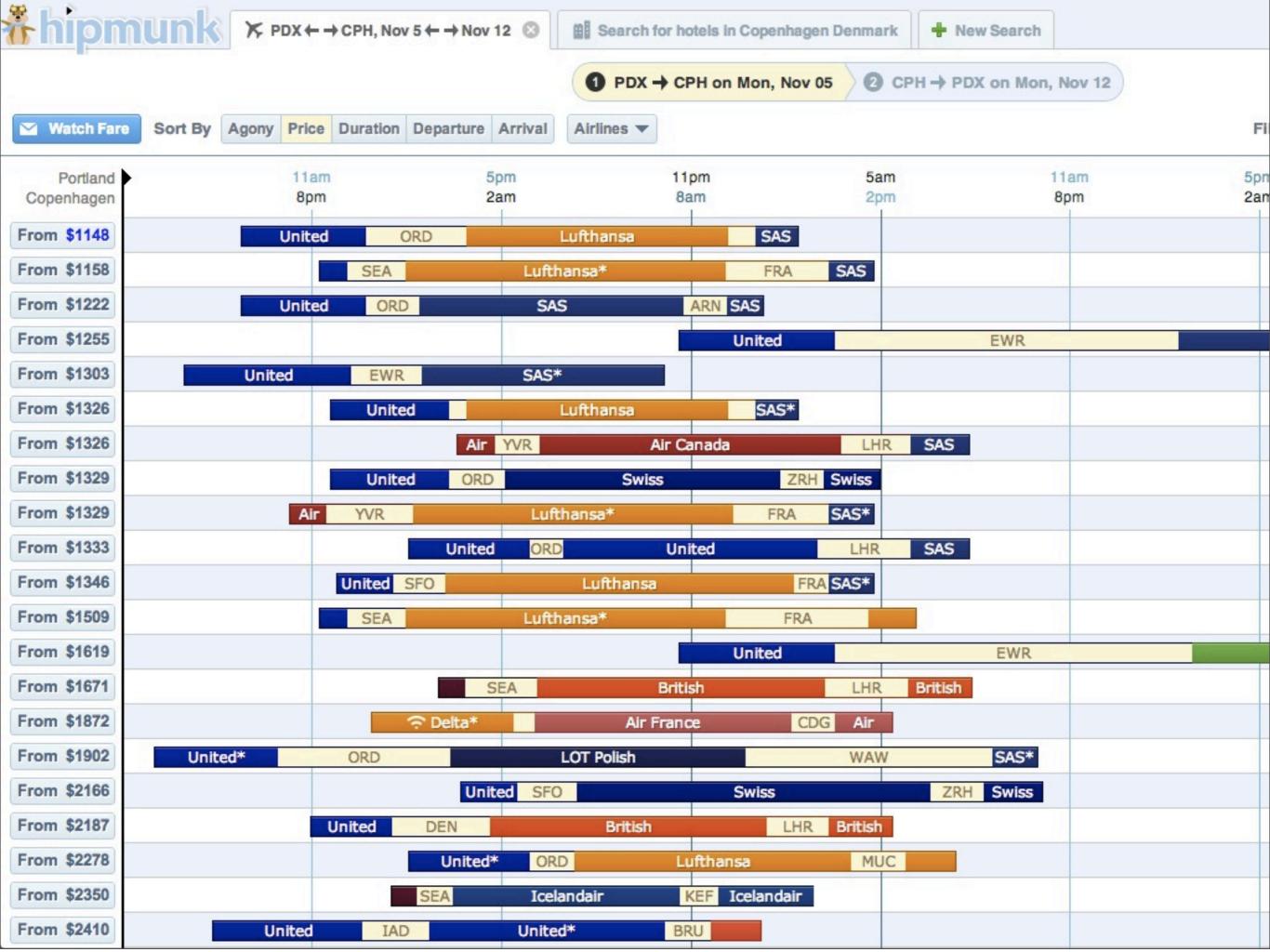


*hip n	TURK > PDX ← → CPH, Nov 5 ← → Nov 12 ② Search for hotels in Copenhagen Denmark + New Search
	● PDX → CPH on Mon, Nov 05 PDX on Mon, Nov 12
🖬 Wata	
Por	didn't make
Copenh	
From \$1	this.
From \$1	
From \$1872	CDG Air
From \$1902	United* ORD LOT Polish WAW SAS*
From \$2166	United SFO Swiss ZRH Swiss
From \$2187	United DEN British LHR British
From \$2278	United* ORD Lufthansa MUC
From \$2350 From \$2410	SEA Icelandair KEF Icelandair
Saturday, Novom	United IAD United* BRU



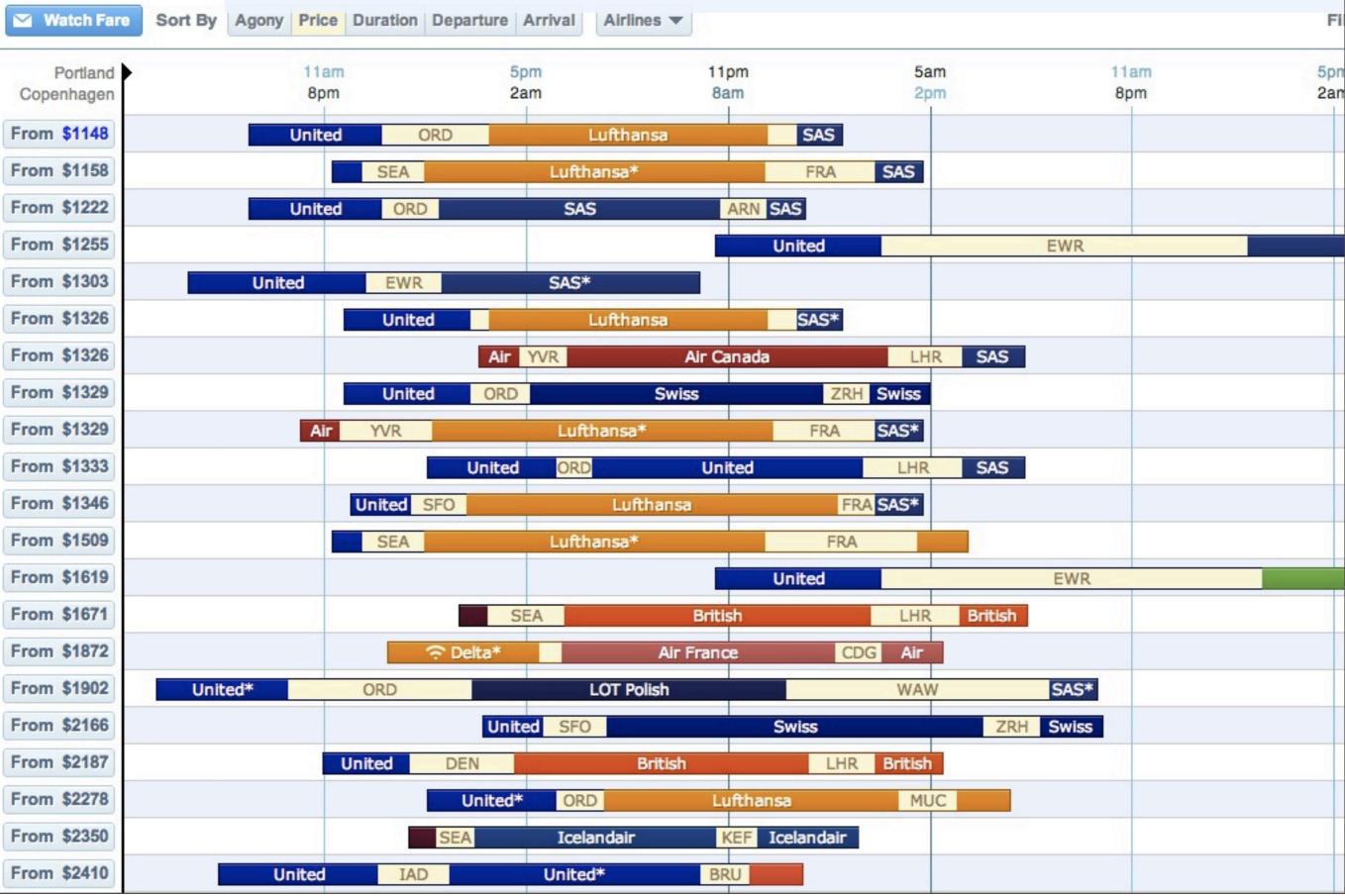
(theory)

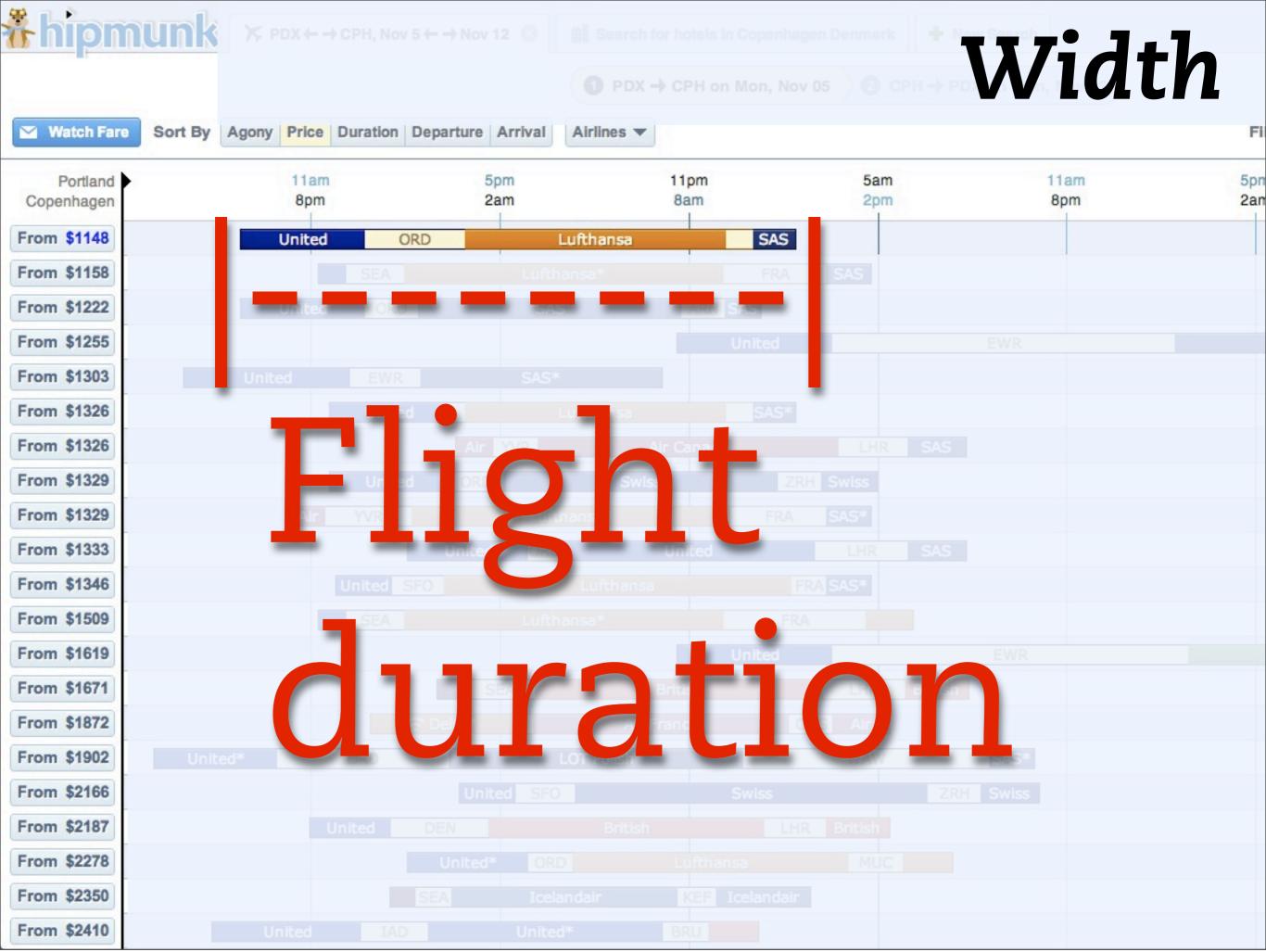


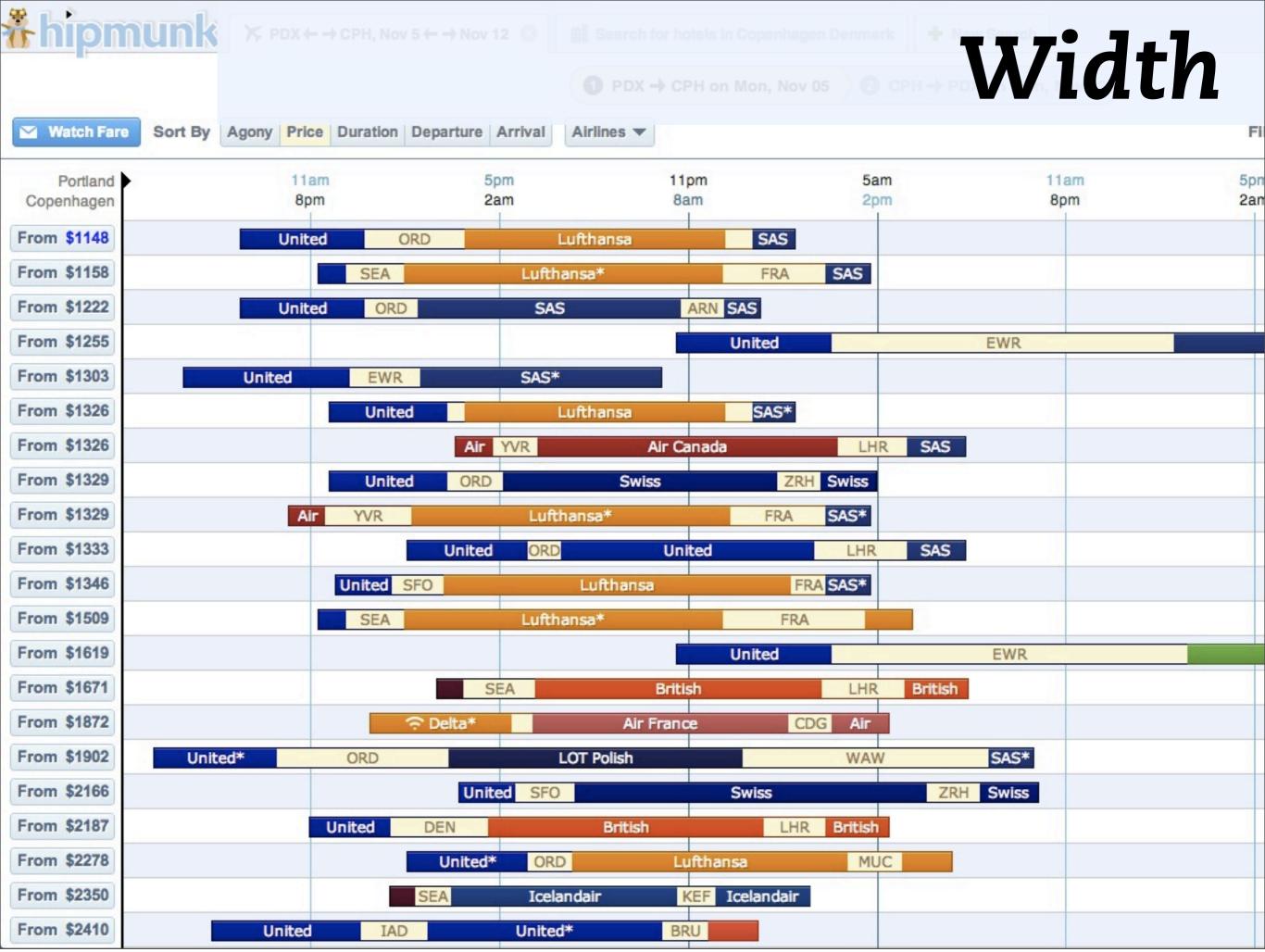




2D Position



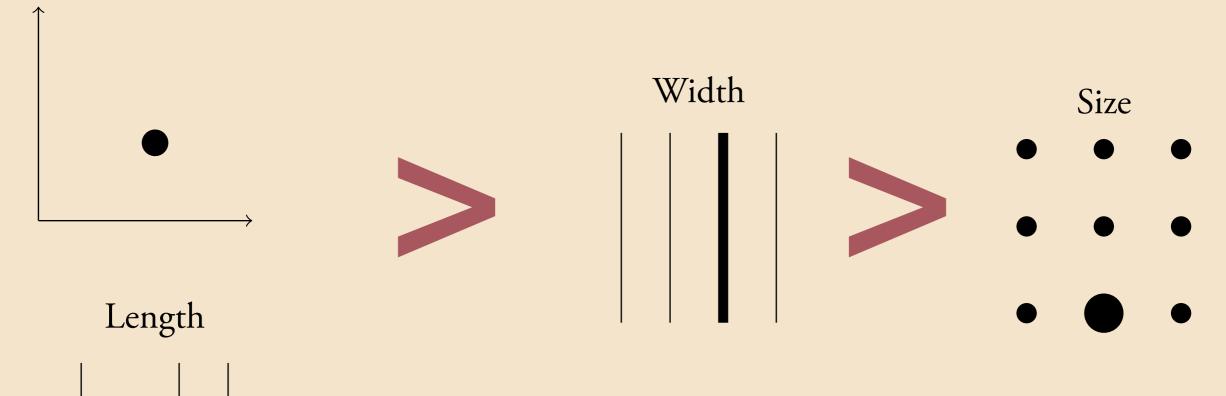




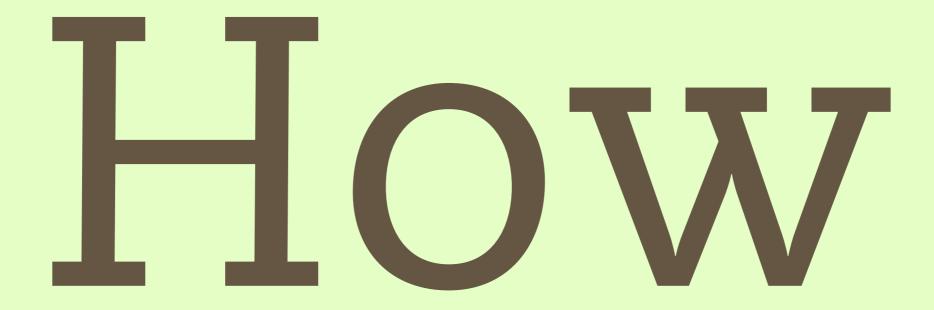
Lessons

Some aesthetics are better than others

2-D position





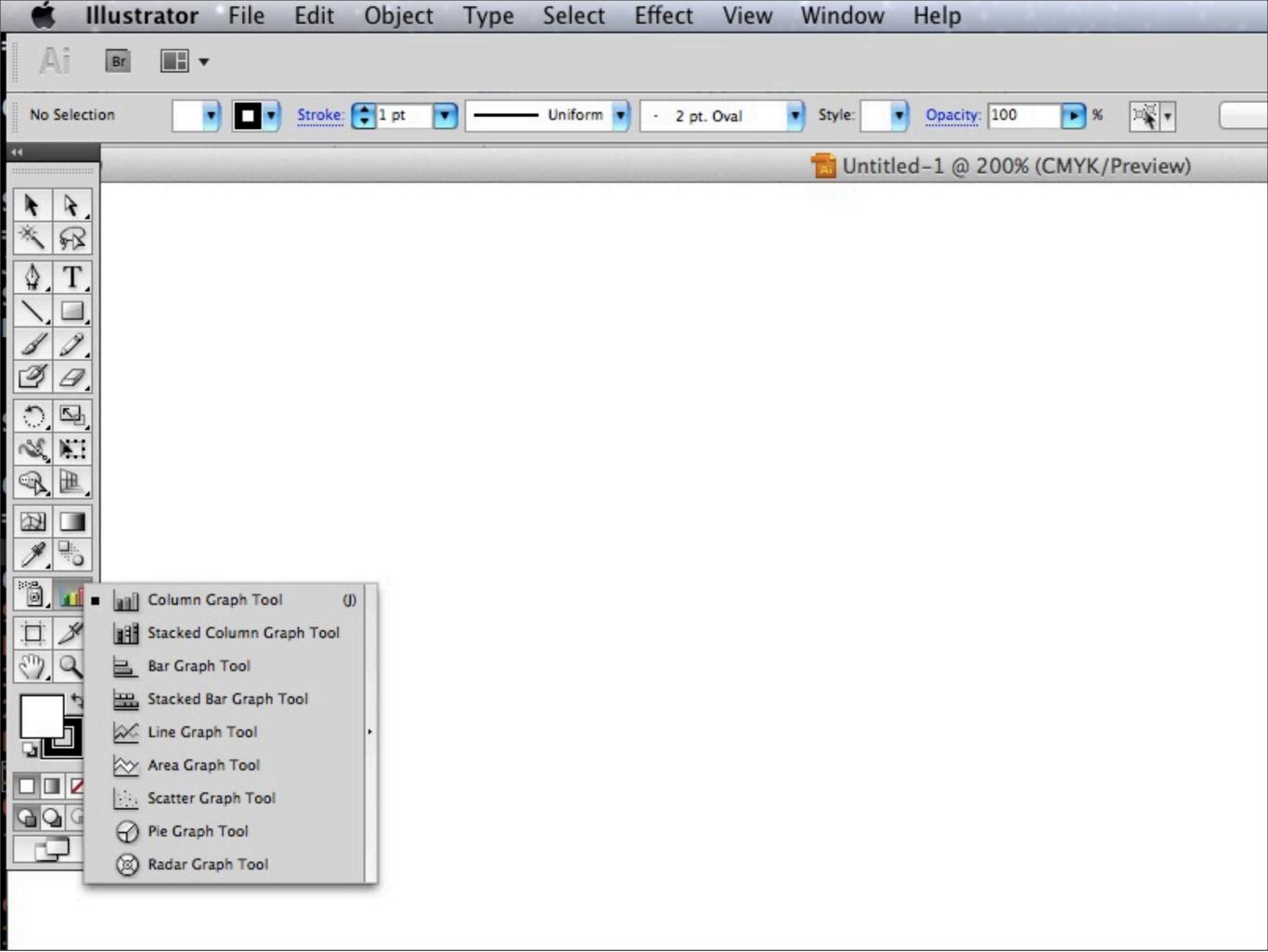


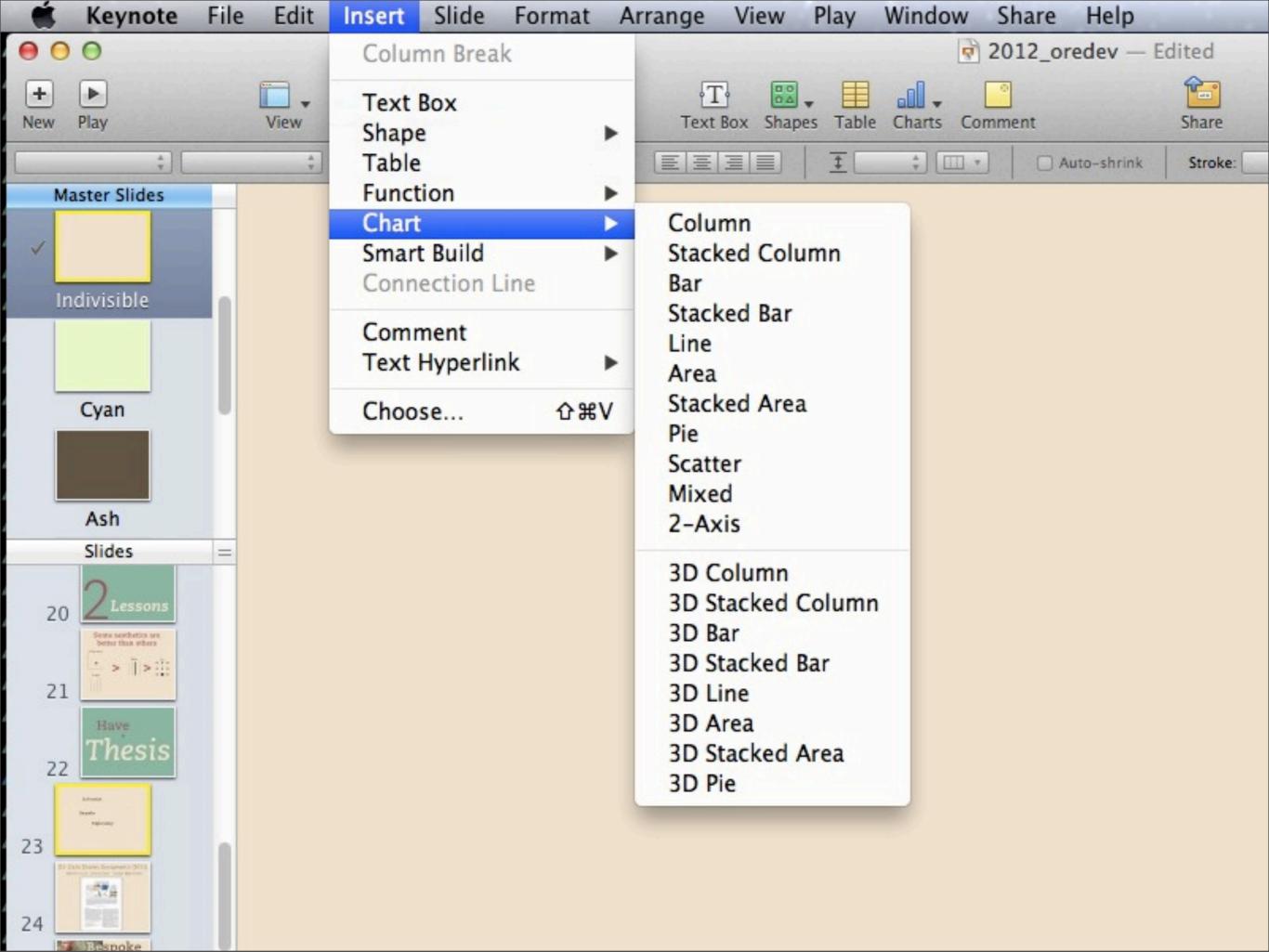
(practice)

Off the rack



		🤊 - 🥲 heet 🛛 I	÷	Pa	ge Layout	Form	ulas Data	a Re	Sales Da		Microsoft Exce	1						-	□ X Ø
1000	tTable Table		Colum			Bar Charts	Area X (Scat	Y C	Dther harts *	re I Gra	GX Shapes aphic *	Kan be a constraint of the second sec	A Text Box	Header & Footer	WordArt Ta	Signature Line *	Object 9	Ω Symbol	
	Н	18	-	0	f _x														×
4	A	В			С	D	E		F	G	Н	I		J	K	L		Μ	
1 2 3		Total S	ales	by S	itore		Total Sa	es bv	Region										
4		Store	and the second second second second	Sale	and the second		Region	Sales		-	1	OTAL	SA	LES B	YCA	TEGO	DRY		
5		Store 1		S	313,765		West	Øs	1,718,258										
6		Store 2		S	107,160		South	Øs	534,389				-						
7		Store 3		S	351,751		Midwest	Os	1,009,268							/			
8		Store 4		S	131,047		East	O S	900,431	-									
9		Store 5		S	252,136		Total	\$	4,162,346										
10		Store 6		S	167,462									1					
11		Store 7		S	210,073														
12	-	Store 8		S	308,092		Total Sa	les by	Category										-
13		Store 9		S	97,492		Category												
14		Store 10		S	393,484		Automotive	and the second s	86,285										
15		Store 11		S	396,891		Gardening	S	52,048				-						
16		Store 12		S	151,168		Electronics	\$	83,026										
17	_	Store 13		S	251,390		Jewelry	S	93,035										
18		Store 14		S	392,776		Sporting	S	50,016					-	-				
19		Store 15		S	259,654		Houseware	S	19,149			Automo	otive	Garde	ening 🛛	Electron	ics		
20		Store 16		S	225,184		Books	S	42,247										
21	-	Store 17		S	335,785		Games	S	18,420			Jewelry	/	Sportin	ng 🖬	Housew	ares		
22				-				-				Books		Game	S				
23								-								19			
24																			
25 26	-																		
		Cummer		-	87 A			_		-								_	
Read	▶ ▶ [y	Summa	y Z D	ata] 100% 🚍		v	-⊕i





Saturday, November 10, 12

Bespoke

Photo by **Scott Schuman**, The Sartorialist

Saturday, November 10, 12

8

.

1

D3: Data Driven Documents (2011) Mike Bostock Jeffrey Heer Vadim Ogievetsky



Fig. 1. Interactive visualizations built with D3, running inside Google Chrome. From left to right: calendar view, chord diagram, choro pleth map, hierarchical edge bundling, scatterplot matrix, grouped & stacked bars, force-directed graph clusters, Voronoi tessellation.

Abstract—Data-Driven Documents (D3) is a novel representation-transparent approach to visualization for the web. Rather than hide the underlying scenegraph within a tookit-specific abstraction, D3 enables direct inspection and manipulation of a native representation: the standard document object model (DOM). With D3, designers selectively bind input data to arbitrary document elements, applying dynamic transforms to both generate and modify content. We show how representational transparency improves expressiveness and better integrates with developer tools then prior approaches, while offering comparable notational efficiency and retaining powerful declarative components. Immediate evaluation of operators further simplifies debugging and allows iterative development Additionally, we demonstrate how D3 transforms naturally enable animation and interaction with dramatic performance improvements over intermediate representations

Index Terms-Information visualization, user interfaces, tookits, 2D graphics,

1 INTRODUCTION

When building visualizations, designers often employ multiple tools it incurs a high opportunity cost: it ignores developers' knowledge of simultaneously. This is particularly true on the web, where interactive standards, and the tools and resources that augment these standards. visualizations combine varied technologies: HTML for page content, CSS for aesthetics, JavaScript for interaction, SVO for vector graphics, and so on. One of the great successes of the web as a platform is the (mostly) seamless cooperation of such technologies, enabled by a shared representation of the page called the document object model (DOM). The DOM exposes the hierarchical structure of page content, such as paragraph and table elements, allowing reference and manipulation. In addition to programming interfaces, modern browsers include powerful graphical tools for developers that display the element tree, reveal inherited style values, and debug interactive scripts.

Unfortunately, this blissful interoperability is typically lost with visualization toolkits due to encapsulation of the DOM with more specialized forms. Rather than empowering direct manipulation of the existing model, such toolkits [2, 9, 18] supplant it with custom scene-may be supported by the underlying representations, they may not be graph abstractions. This approach may provide substantial gains in efficiency-reducing the effort required to specify a visualization-but

- · The authors are with the Computer Science Department of Stanford University Stanford, CA 94305. Email: (mbostock.vad,jheer)@stanford.edu.

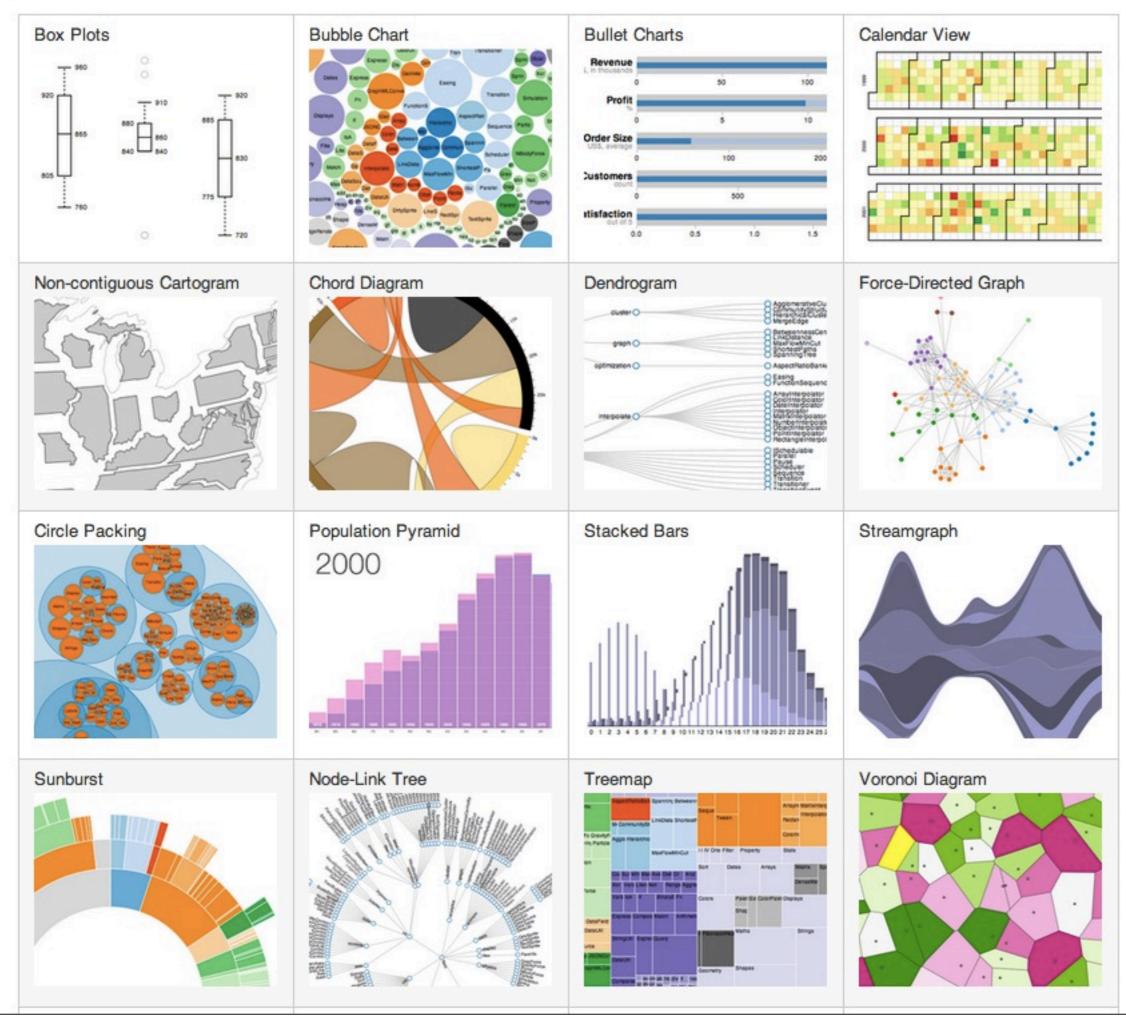
Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011. For information on obtaining reprints of this article, please send email to: tvcg@computer.org

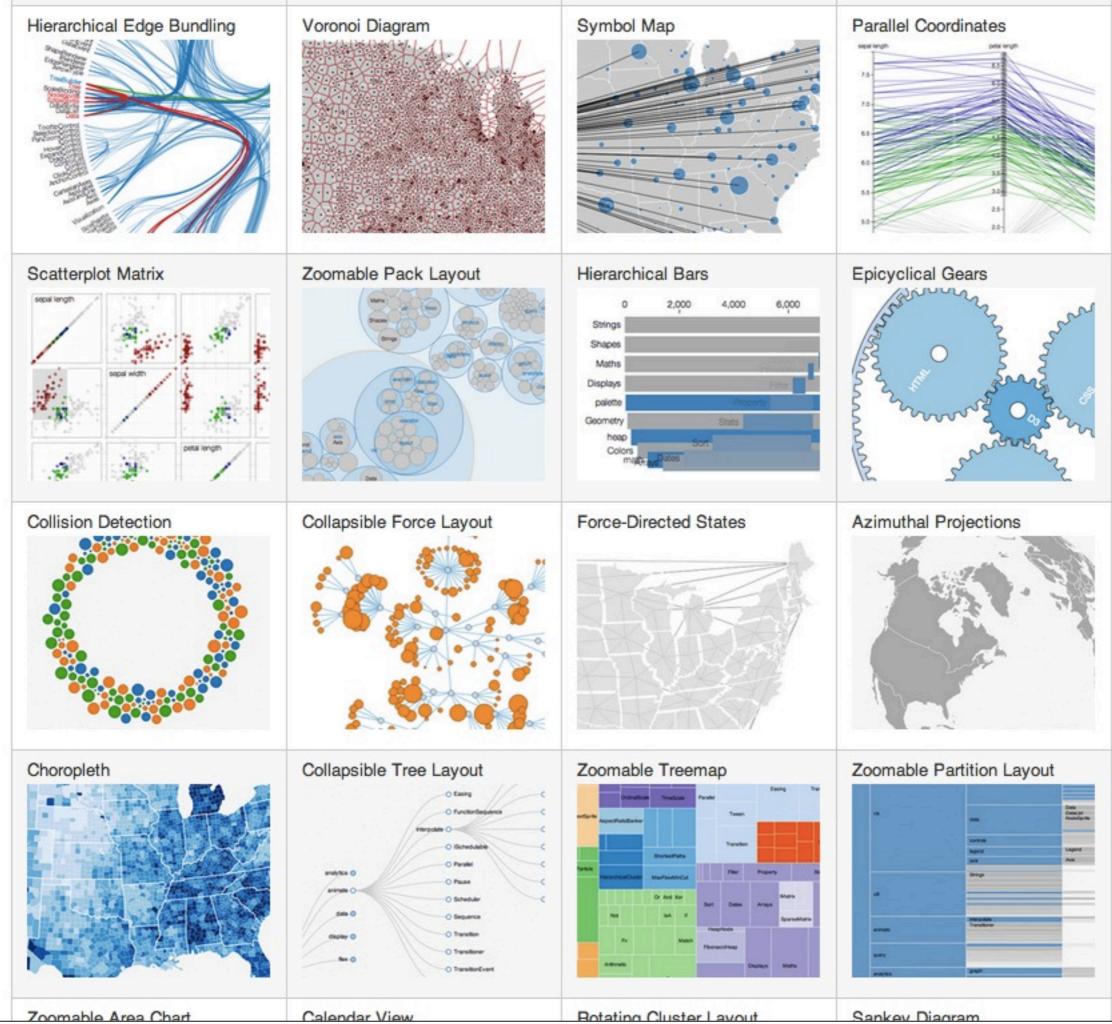
The resulting cost to accessibility-the difficulty of learning the representation-may trump efficiency gains, at least for new users. Scarcity of documentation and ineffectual debugging exacerbate the problem, impeding users from gaining deeper understanding of toolkit abstractions and limiting the toolkit's potential. Systems with intermediate scenegraph abstractions and delayed property evaluation can be particularly difficult to debug: internal structures are exposed only when errors arise, often at unexpected times.

Furthermore, intermediate representations may diminish expressiveness-the diversity of possible visualizations-and introduce substantial runtime overhead. Certain tasks that could be offloaded to a more suitable tool, such as specifying fonts via CSS, may be stymied exposed by the toolkit. Even if extensibility is available as a means for greater expression, it requires in-depth knowledge of toolkit internals and poses a substantial barrier to the average user.

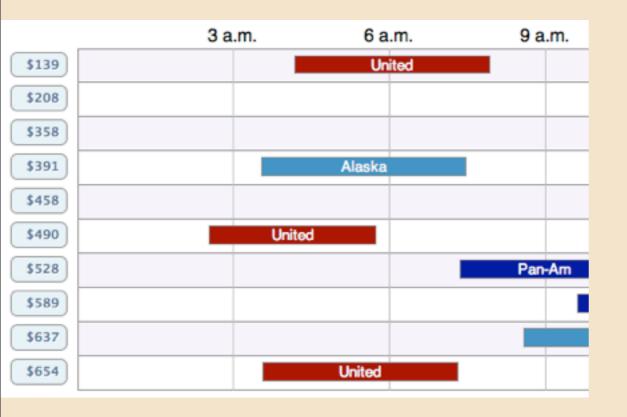
Our awareness of these issues comes in part from thousands of user observations over the two years since releasing Protovis [2], despite our attempt to balance expressiveness, efficiency and accessibility. We now refine these three goals with specific objectives:

Compatibility. Tools do not exist in isolation, but within an ecosystem of related components. Technology reuse utilizes prior knowledge and reference materials, improving accessibility. Offloading a subset of tasks to specialized tools can improve efficiency, avoiding the generality and complexity of a monolithic approach. And, full access to





Bespoke



Bespoke

	3 a	.m. 6a	.m. 9a	a.m.
\$139		Un	ited	
\$208				
\$358				
\$391		Alaska		
\$458				
\$490		United		
\$528			Par	-Am
\$589				
\$637				
\$654		United		



	3 a	.m. 6a	.m. 9	a.m.
\$139		Un	ited	
\$208				
\$358				
\$391		Alaska		
\$458				
\$490		United		
\$528			Pa	in-Am
\$589				
\$637				
\$654		United		

```
#flig
                                       #axis
                                          .ti
                                                  ic : ab ol t
                                              si
            3 a.m.
                       6 a.m.
                                 9 a.m.
                                                                                      ransi
                                            height: 100%
$139
                       United
                                                                       ease-out
                                            .label
$208
                                                                        .price
                                              display: block
$358
                                                                         position: absolu
                                              position: relative
                     Alaska
$391
                                                                         text-align: righ
                                              top: -1.3em
$458
                                                                         width: 50px
                                              left: -2.5em
$490
                United
                                                                         height: 80%
                                              width: 5em
                                 Pan-Am
$528
                                                                         top: 10%
                                              text-align: center
$589
                                                                          left: -60px
                                              font-family: sans-serif
$637
                                            .grid-line
$654
                     United
                                                                        .flight
                                              position: absolute
                                                                         position: absolu
                                              width: 1px
                                                                         height: 60%
(map flight-data
                                              height: 100%
  (fn [{:keys [price carrier depart arrive]}]; 0px
                                                                         top: 20%
                                              background-color: hsl(0,05,75, align: cent
      [:div.row
                                                                         border: 1px soli
        [:button.price (str "$" price)]
                                              z-index: 1
        [:div.flight
                                                                          span
                                                                            color: white
           {:style {:left (time-scale depart)
                                                                            font-size: 0.8
                     :width (time-scale (- arrive depart))}
                                                                            font-family: s
            :carrier carrier}
           [:span carrier](hap) [[3 "3 a.m."] [6 "6 a.m."] [9 "9 a.m."]
                                 [12 "Noon"] [15 "3 p.m."] [18 "6 p.m. arrier-col
                                                                       $color)
                                (fn [[t label]]
                                   [:div.tick {:style {:left (time-scale ight[carrier=
                                                                            background-col
                                      [:div.grid-line]
                                      [:span.label label]]))
```

<canvas> Java2D D3.js <svg>

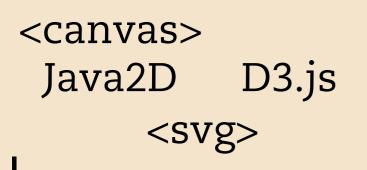
General

(hard to use)

Excel Highcharts

Specific

(not expressive)



Excel Highcharts

General

(hard to use)

Specific

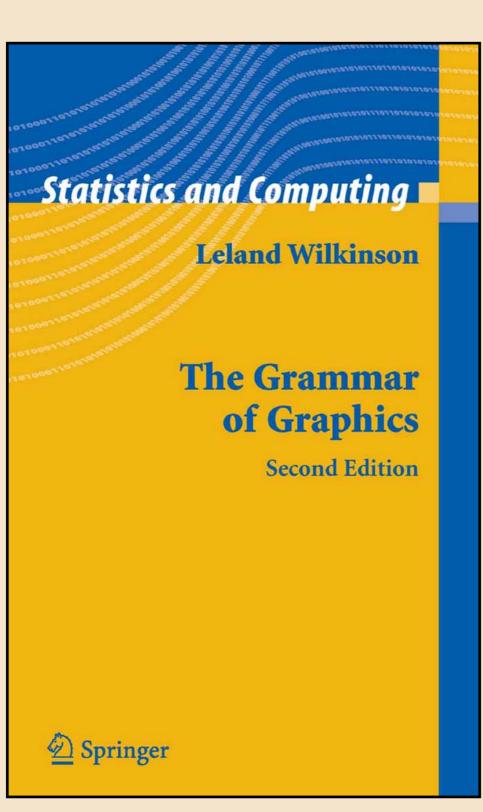
(not expressive)

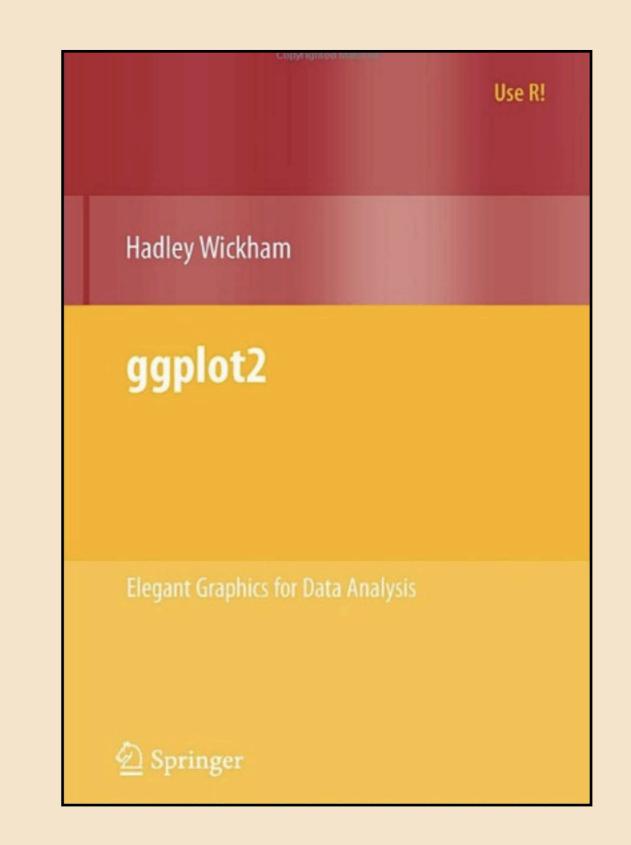


Grammar Graphics

The Grammar of Graphics (1999)

ggplot2 (2005)

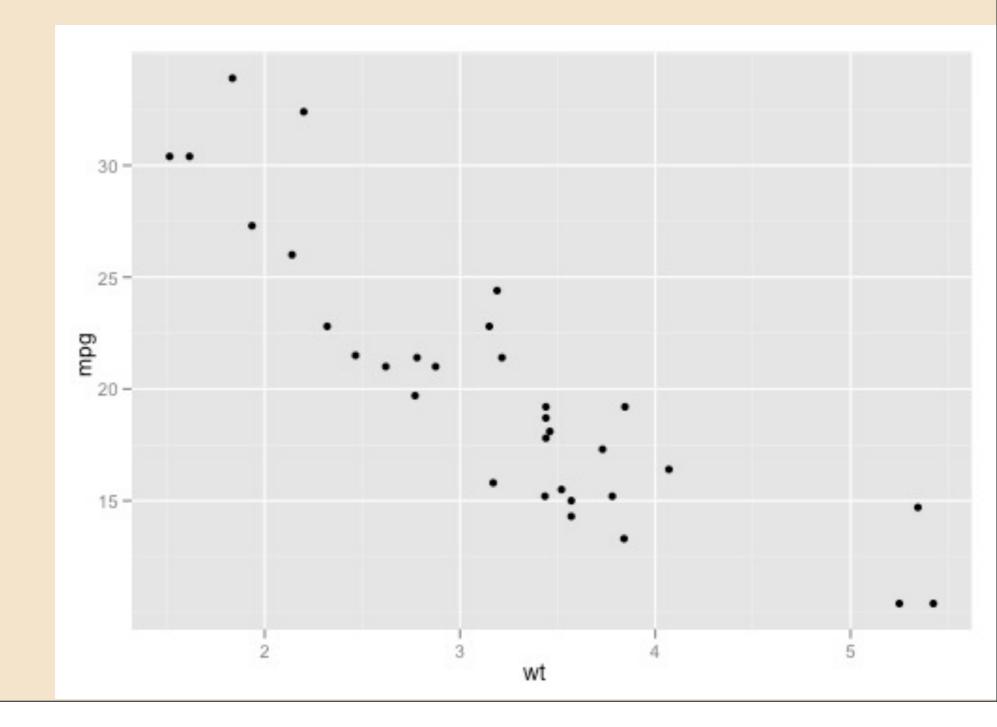




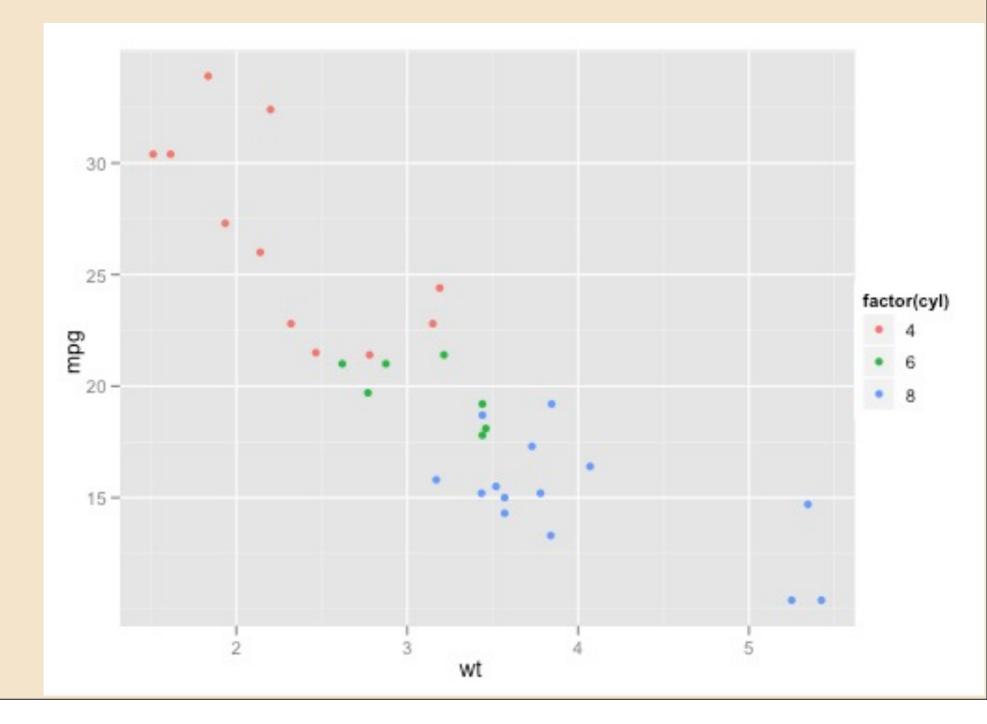
Motor Trend cars dataset

	mpg	cyl	disp	hp	drat	wt	• • •
Mazda RX4	21.0	6	160	110	3.90	2.62	• • •
Mazda RX4 Wag	21.0	6	160	110	3.90	2.88	• • •
Datsun 710	22.8	4	108	93	3.85	2.32	• • •

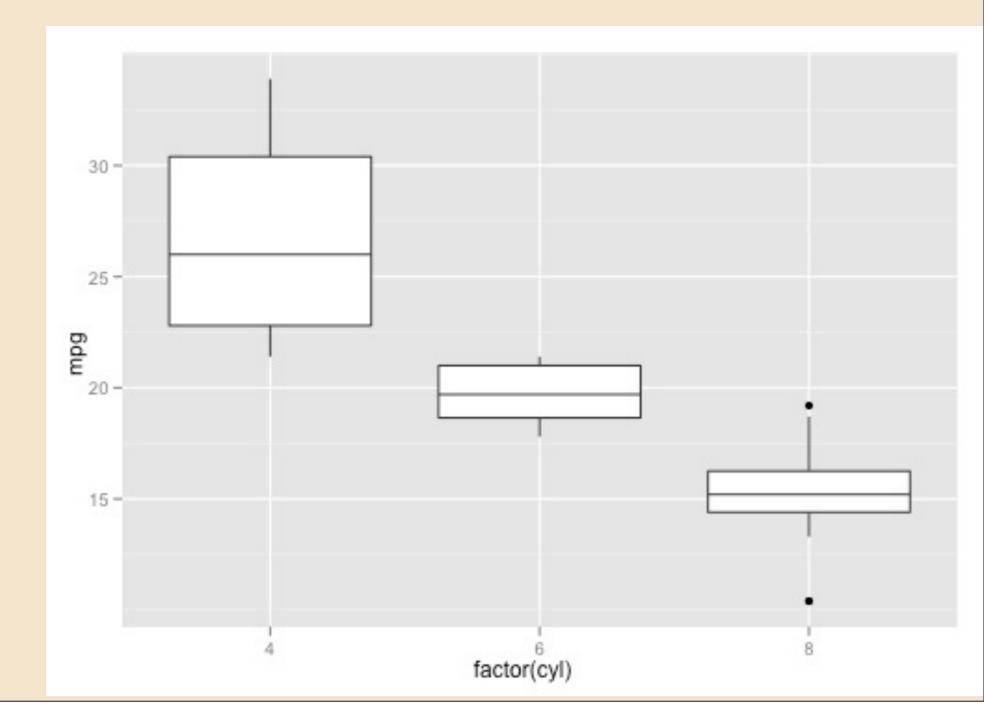
ggplot(mtcars, aes(wt, mpg)) + geom_point()



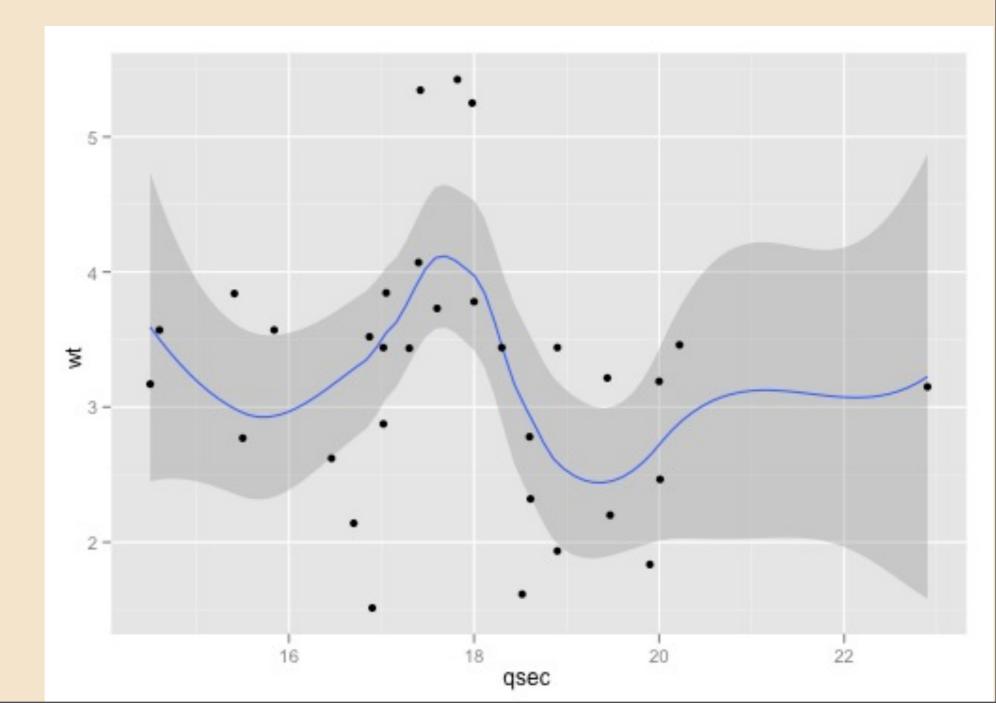
ggplot(mtcars, aes(factor(cyl), mpg))
+ geom_point(aes(colour = factor(cyl)))

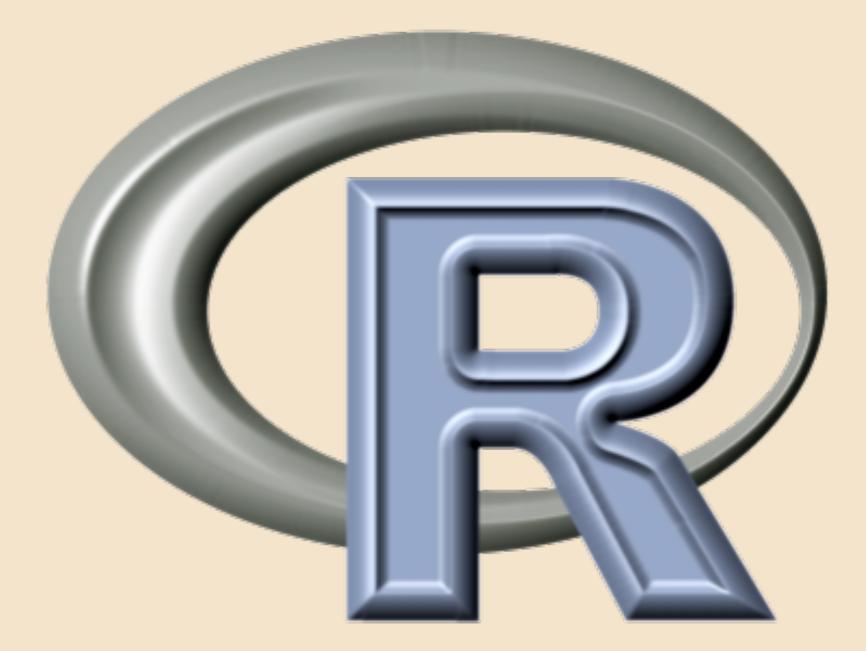


ggplot(mtcars, aes(wt, mpg)) + geom_boxplot()



ggplot(mtcars, aes(qsec, wt))
+ stat_smooth() + geom_point()



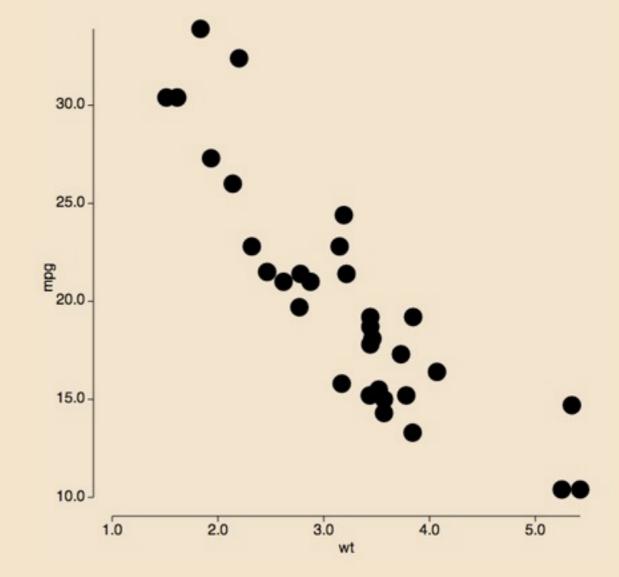






complected

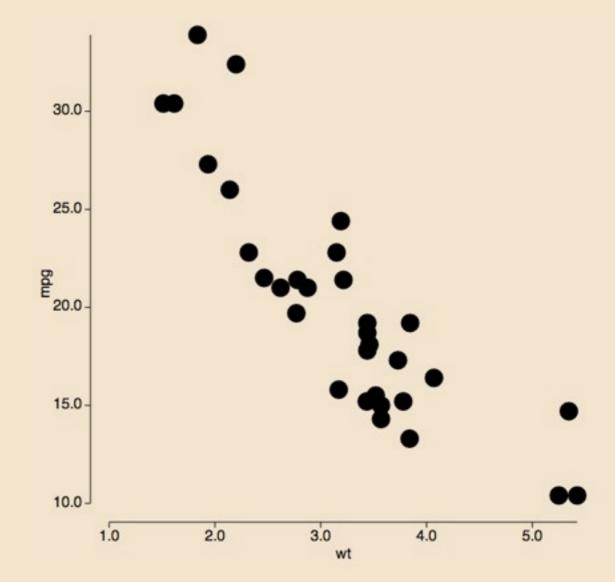




complected

draw_scatterplot(data)

calculate scale extents calculate tick marks curve fitting calculate statistics draw points draw guides draw labels



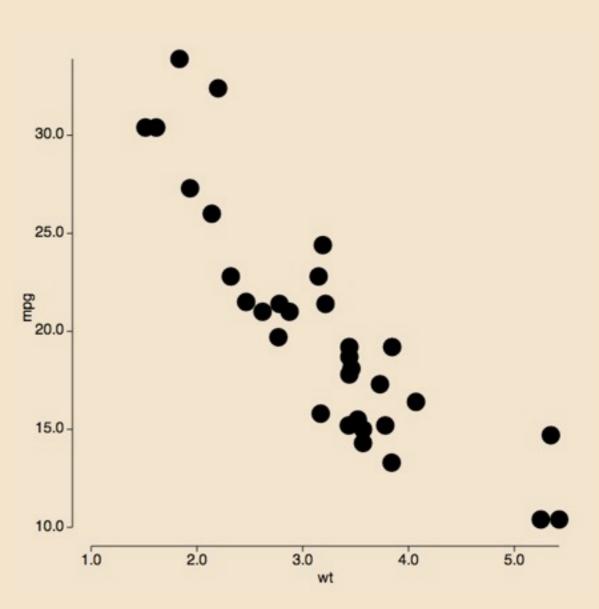
. . .

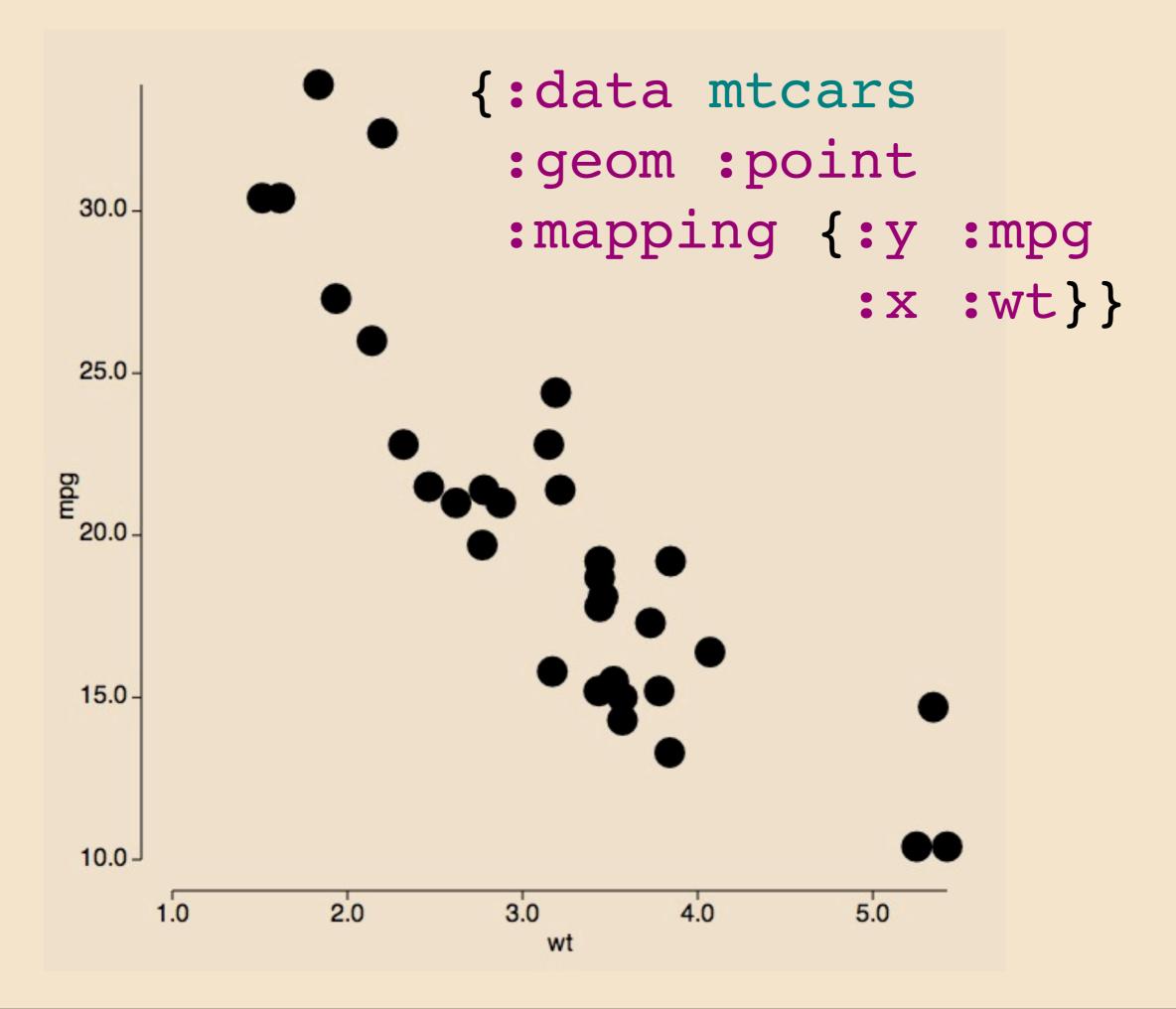


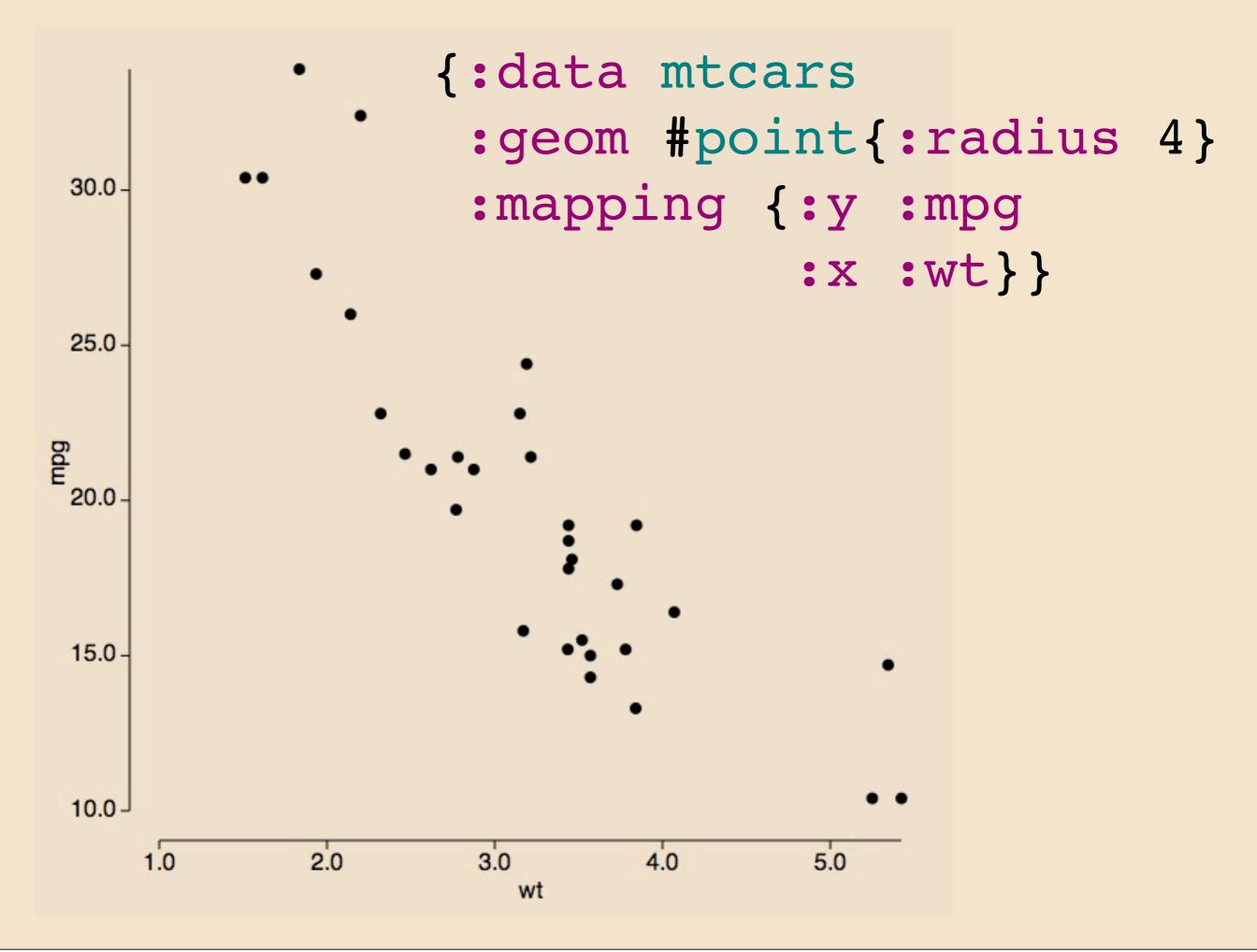
Saturday, November 10, 12

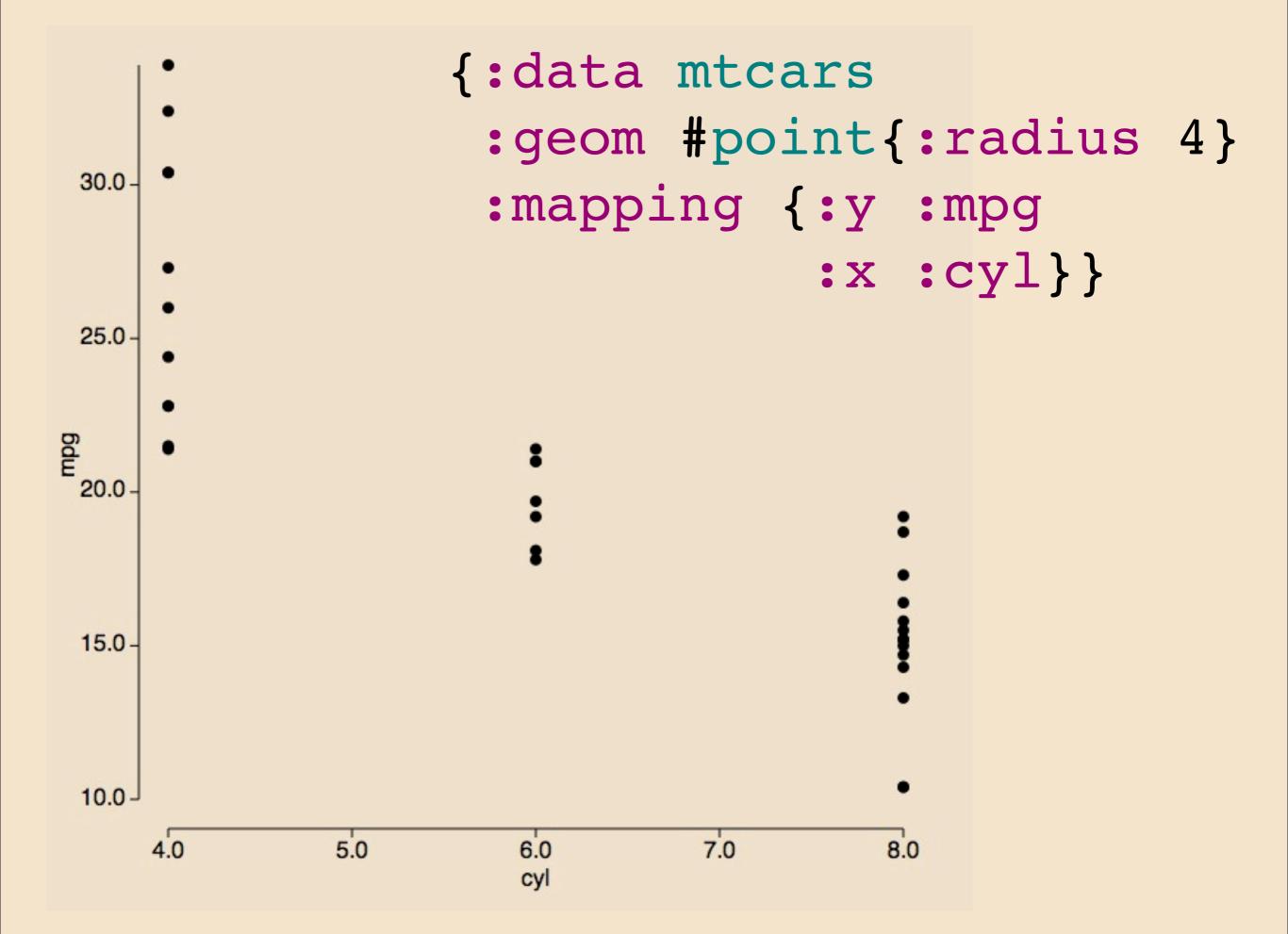
Decomplect

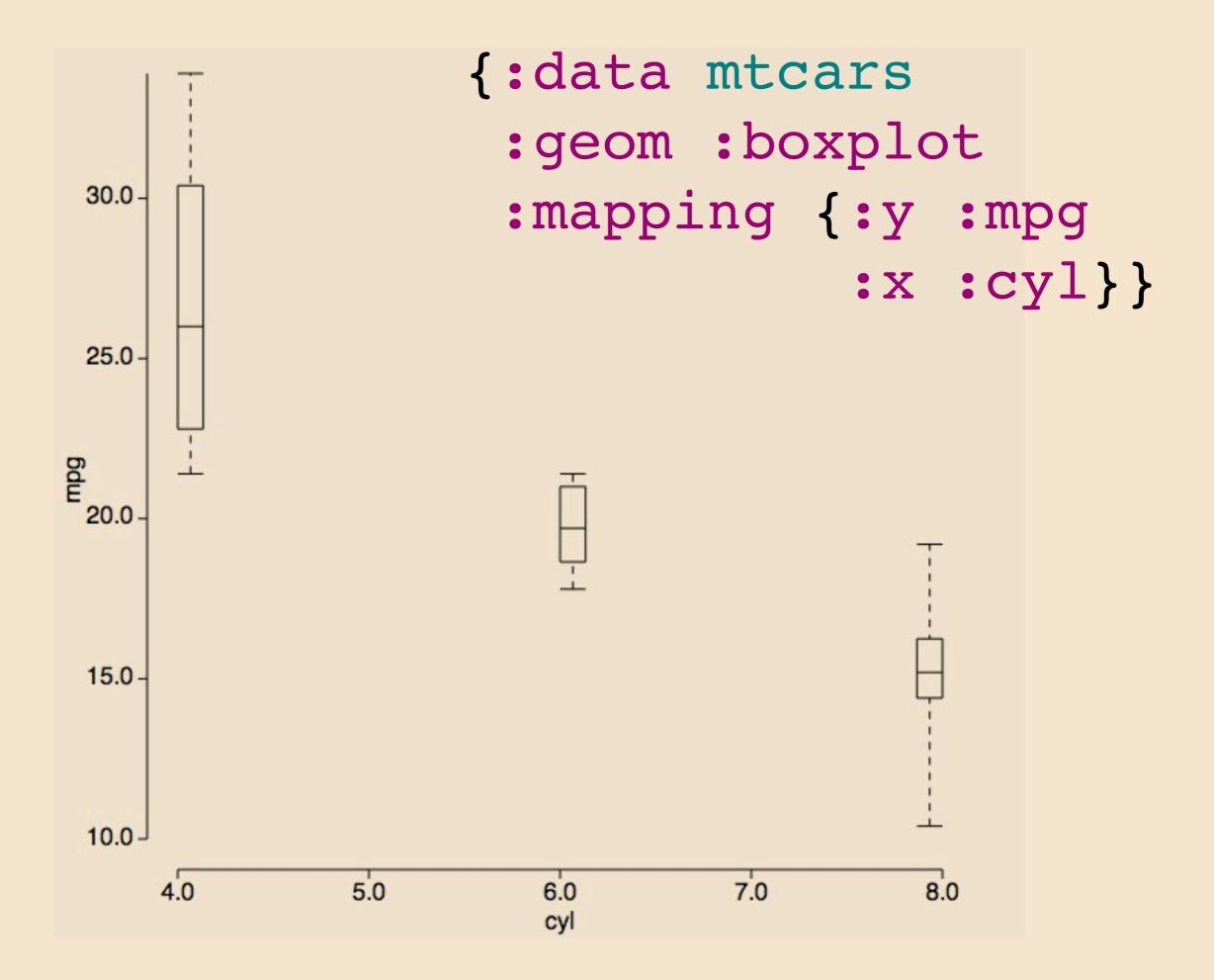
A graphic consists of: Data Geometry Aesthetic mappings **Statistics** Groupings **Scales**













all the things

Saturday, November 10, 12

API Support

- {:data mtcars
 - :mapping {:x :wt :y :mpg}
 - :geom :point}

API Support

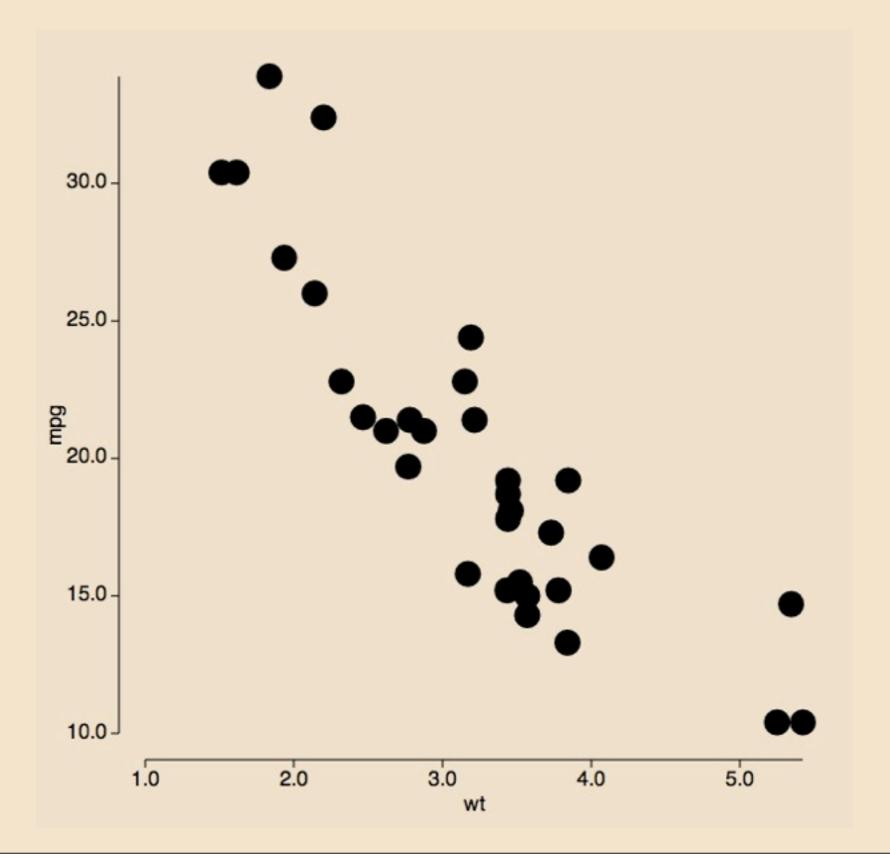
- {:data => mtcars,
 - :mapping => {:x => :wt, :y => :mpg},
 - :geom => :point}

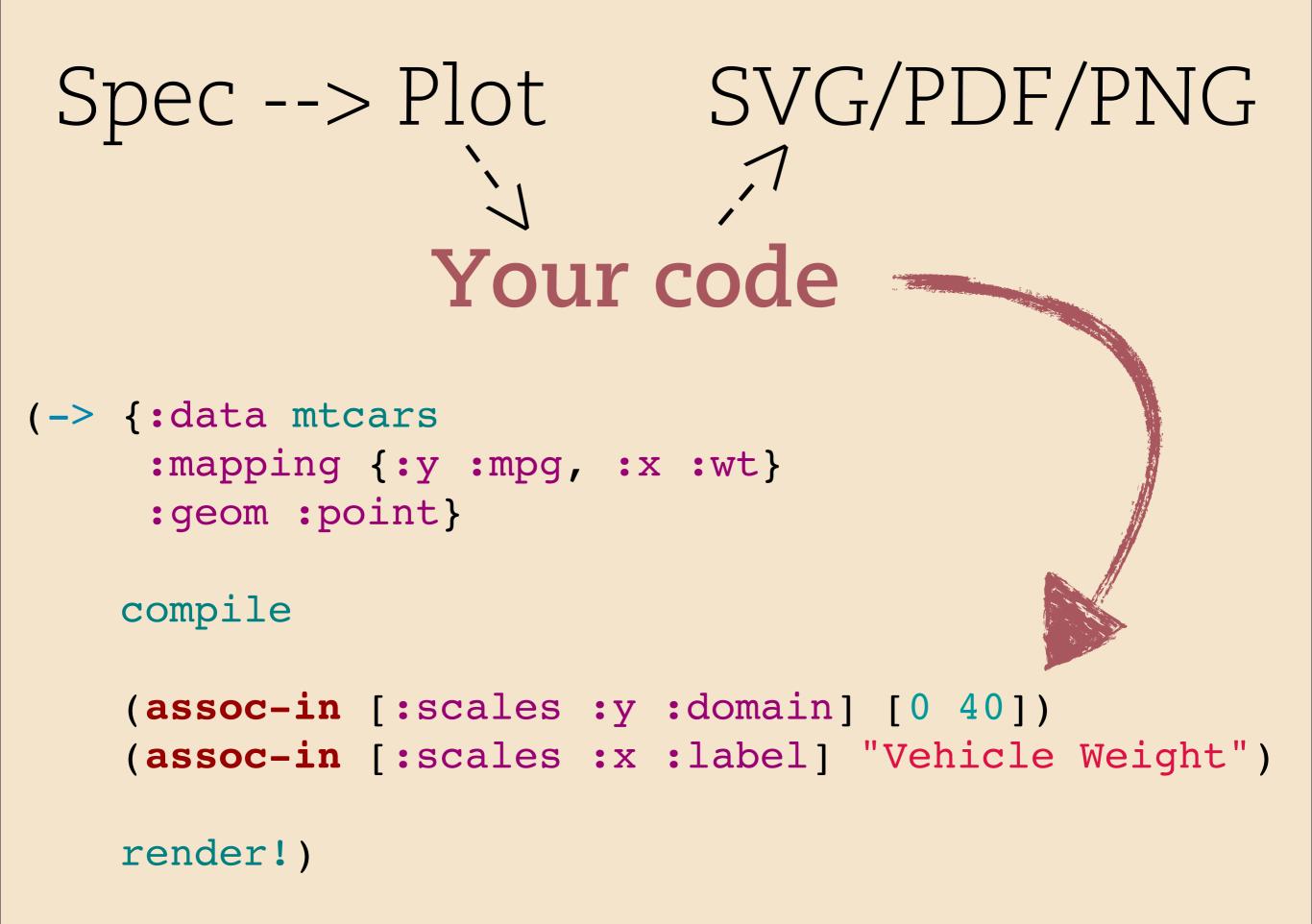
API Support

{"data": mtcars,
 "mapping": {"x": "wt", "y": "mpg"},
 "geom": "point"}

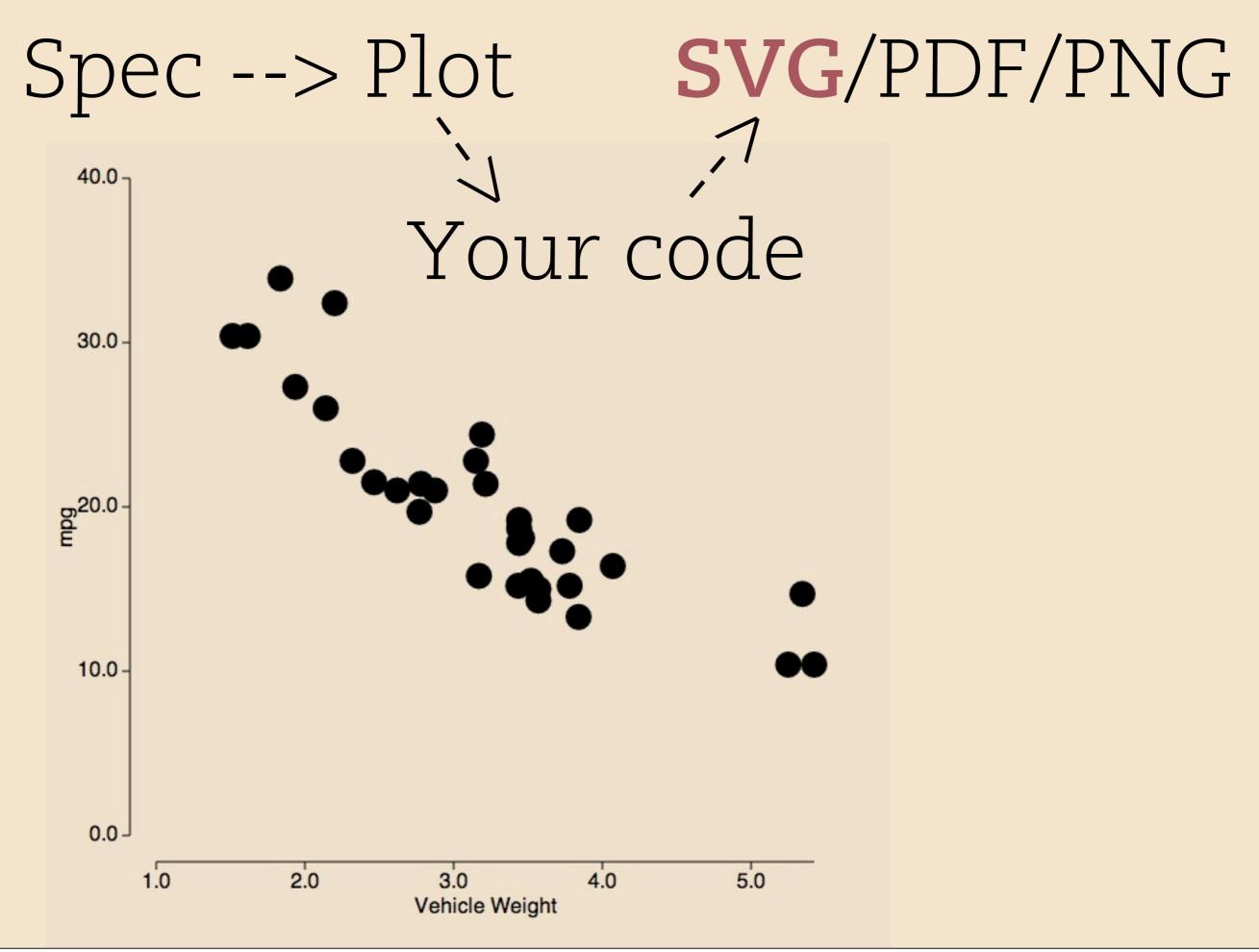
U can haz Programming

{:data mtcars
 :mapping {:x :wt :y :mpg}
 :geom :point}



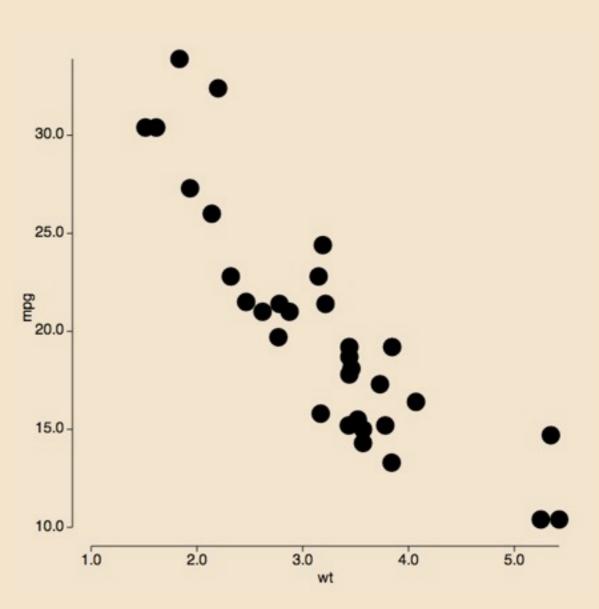


Saturday, November 10, 12



Decomplect

A graphic consists of: Data Geometry Aesthetic mappings **Statistics** Groupings **Scales**



A

grammar lets you



Saturday, November 10, 12



grammar lets you this contract of the second se

Kening Labs @lynaghk