

Data Visualization for Policy Advocacy

David Epstein

Research Associate

Baltimore Neighborhood Indicators Alliance

Jacob France Institute

University of Baltimore

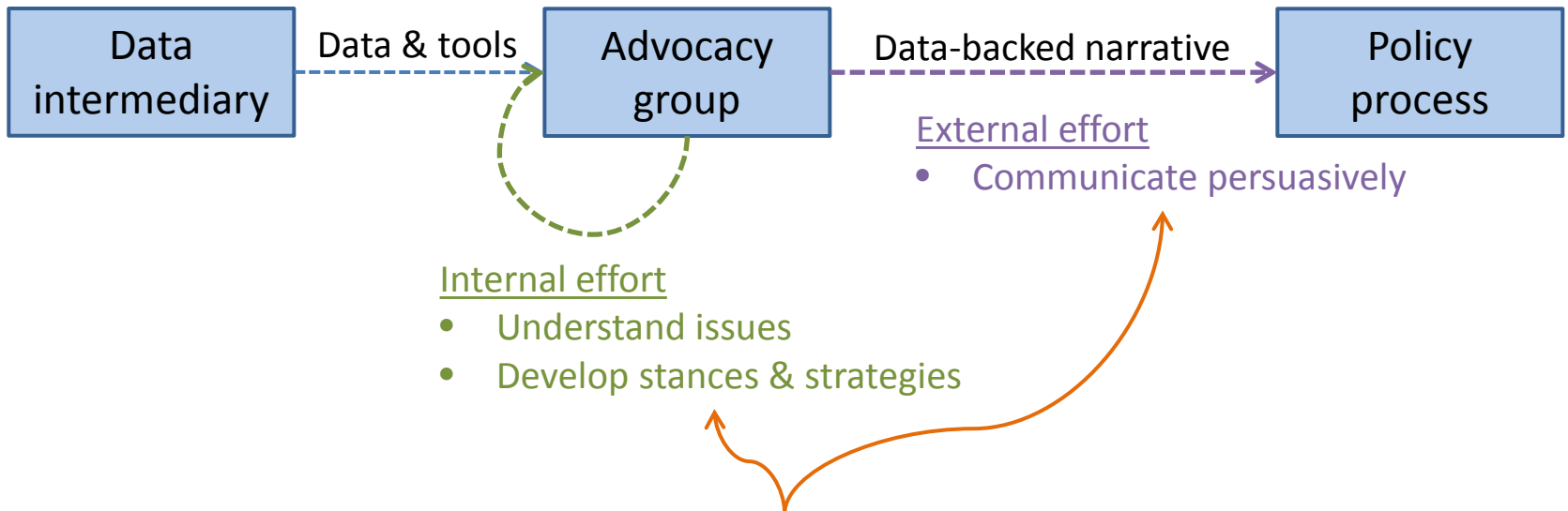
Presentation for the National Low Income Housing Coalition
Housing Policy Conference April 27-30 2014



Data Visualization: The visual representation of
data to improve cognition

Based on a definition by Dr. Eytan Adar, University of Michigan, 2010

Locating data visualization within advocacy

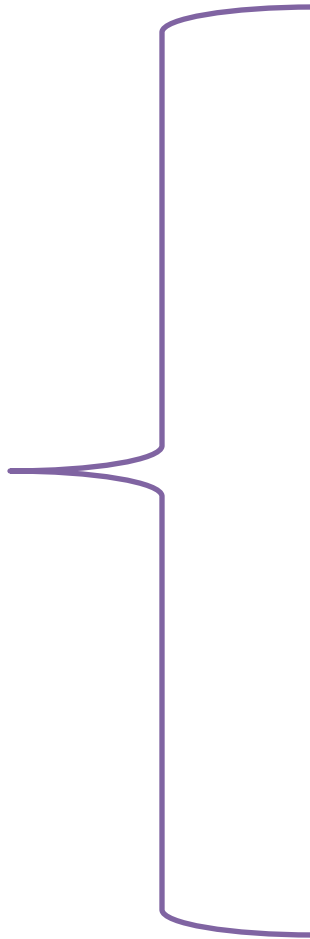


Data visualization: Both internal & external efforts should be informed by an explicit *theory of change* that identifies the target audience of data visualizations (e.g. policy makers, domain experts, the media, or the general public).

Function of Visualization in Advocacy

	Explore	Explain
Purpose	Reveal relationships between variables	Communicate specific points in a narrative
Form	<i>Usually:</i> many rough, static graphics <i>Recently:</i> interactive linked (“brush”) plots	<i>Usually:</i> a few polished, static graphics with text. <i>Recently:</i> a static or interactive infographic.

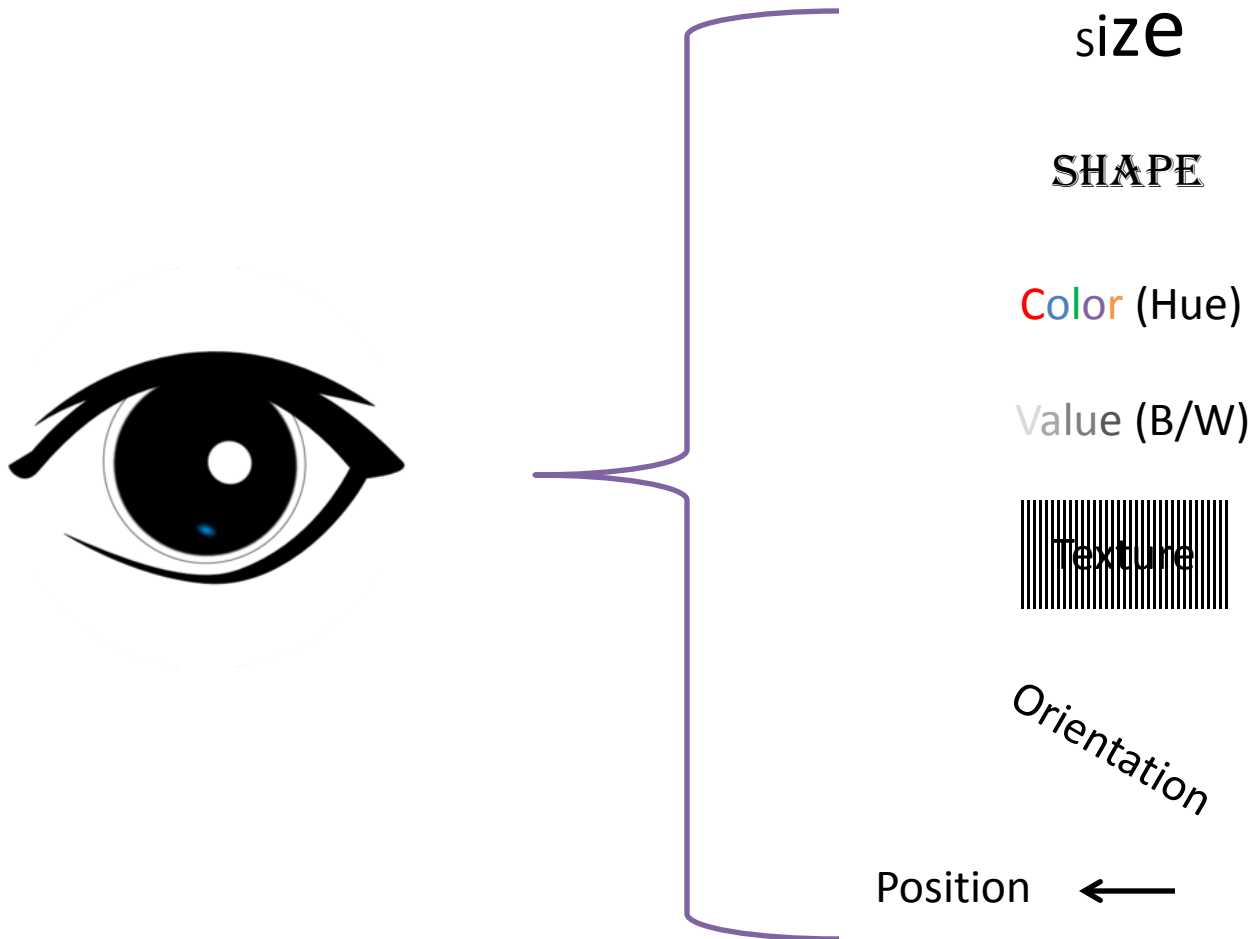
Quick! What's the pattern?



Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

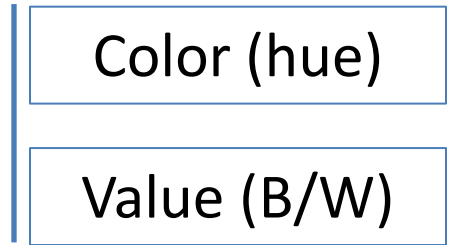
Hypothetical housing data

Human vision generally excels at seeing



Based on Jacques Bertin's "retinal variables" in *Semiology of Graphics* 1983

Does this help?



Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

While pretty, this conditional formatting does **not** improve cognition.

Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of a single categorical (text) variable

Frequency table

Type	Count
Vacant	3
Occupied	7

There are roughly twice as many occupied houses as vacant houses.

Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of a single categorical (text) variable

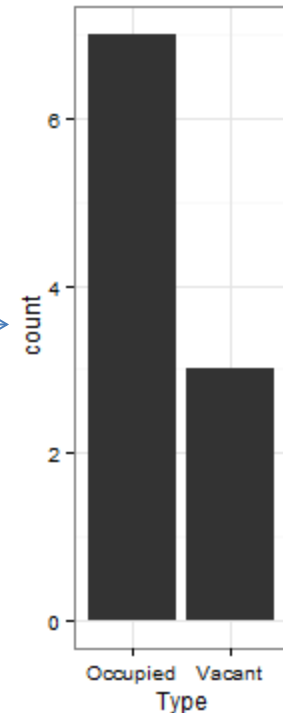
Frequency table

Type	Count
Vacant	3
Occupied	7

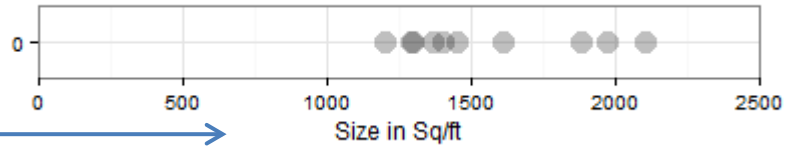
There are roughly twice as many occupied houses as vacant houses.

Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Size

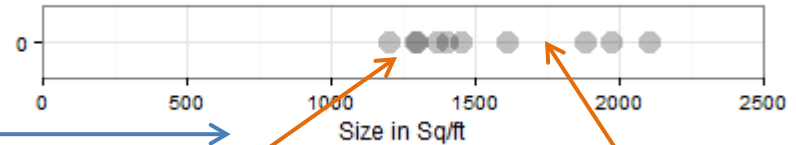


Visualization of single numeric variable



Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of single numeric variable



Many houses
around 1,400 sq/ft
(darker gray).

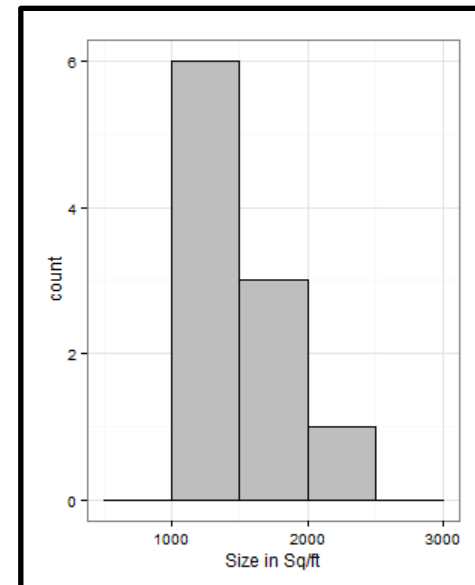
Split between small
and large houses at
around 1,700 sq/ft.

Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of single numeric variable



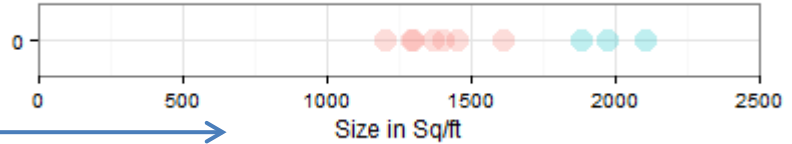
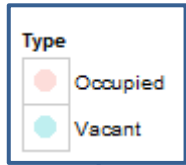
Binning



Careful when setting bin sizes!

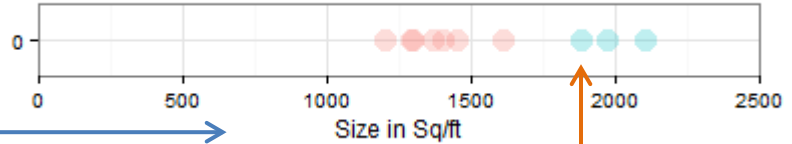
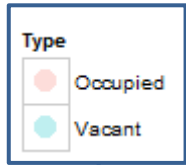
Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of multiple variables



Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

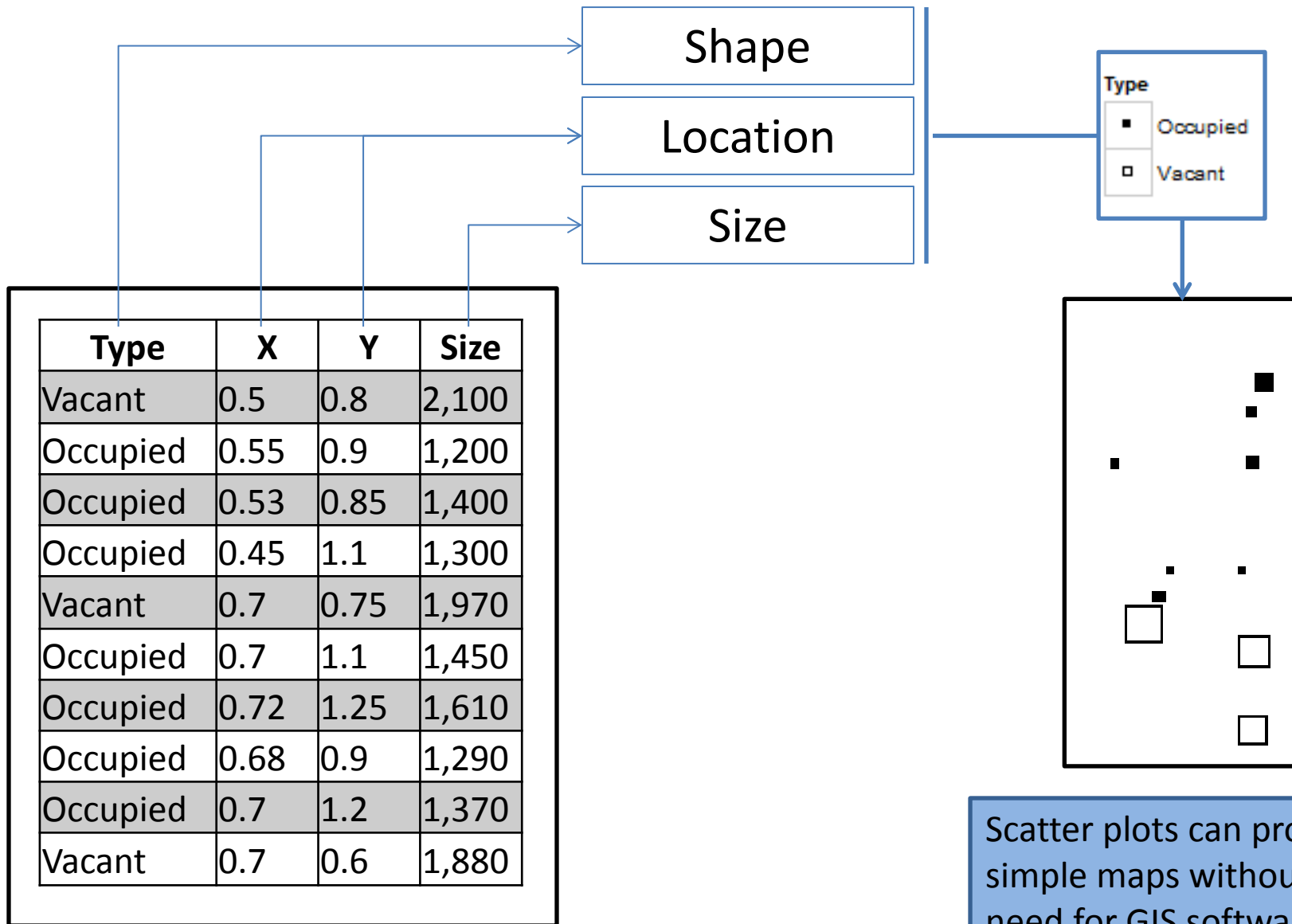
Visualization of multiple variables



Vacant houses are *all* larger than occupied houses.

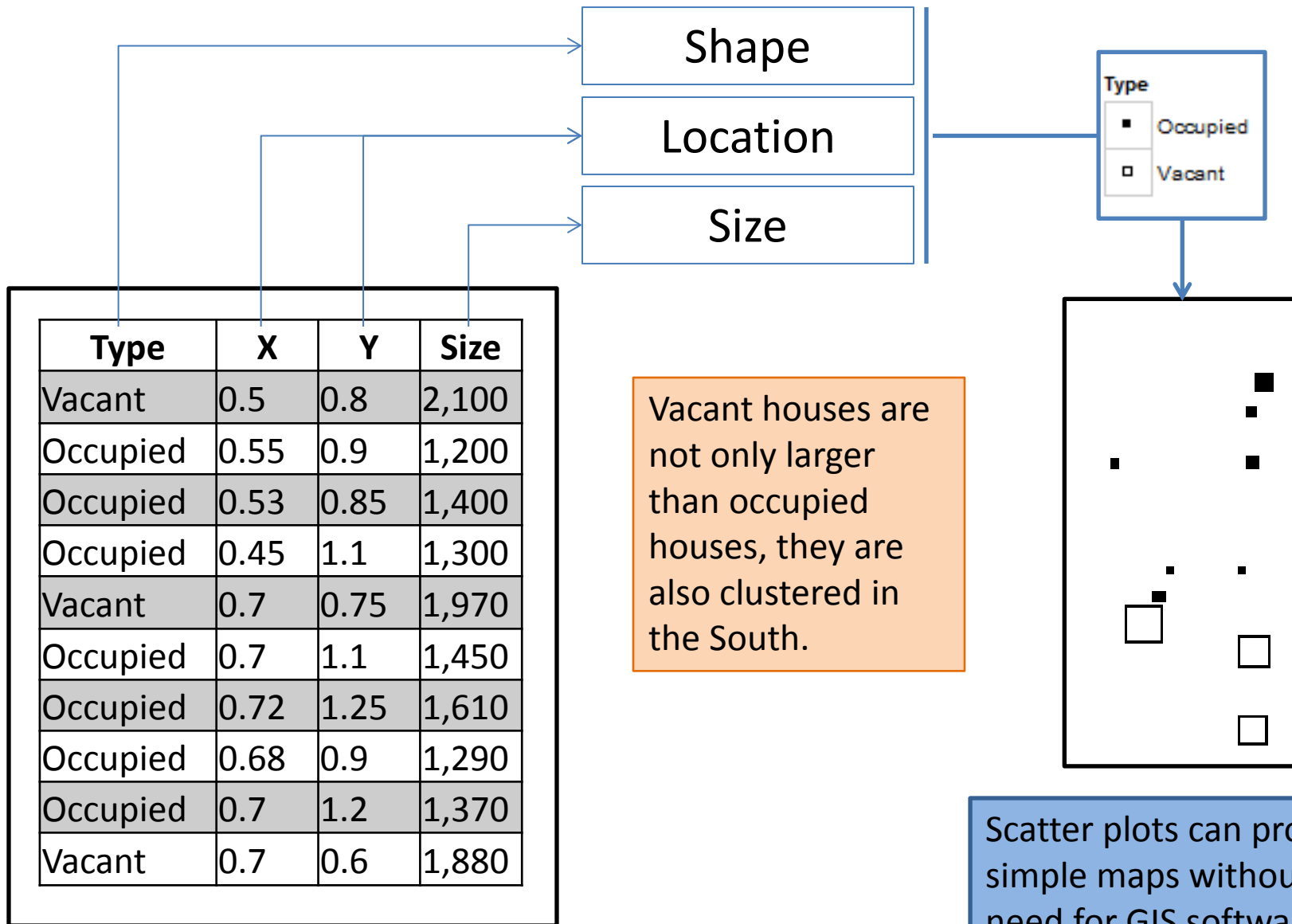
Type	X	Y	Size
Vacant	0.5	0.8	2,100
Occupied	0.55	0.9	1,200
Occupied	0.53	0.85	1,400
Occupied	0.45	1.1	1,300
Vacant	0.7	0.75	1,970
Occupied	0.7	1.1	1,450
Occupied	0.72	1.25	1,610
Occupied	0.68	0.9	1,290
Occupied	0.7	1.2	1,370
Vacant	0.7	0.6	1,880

Visualization of multiple variables



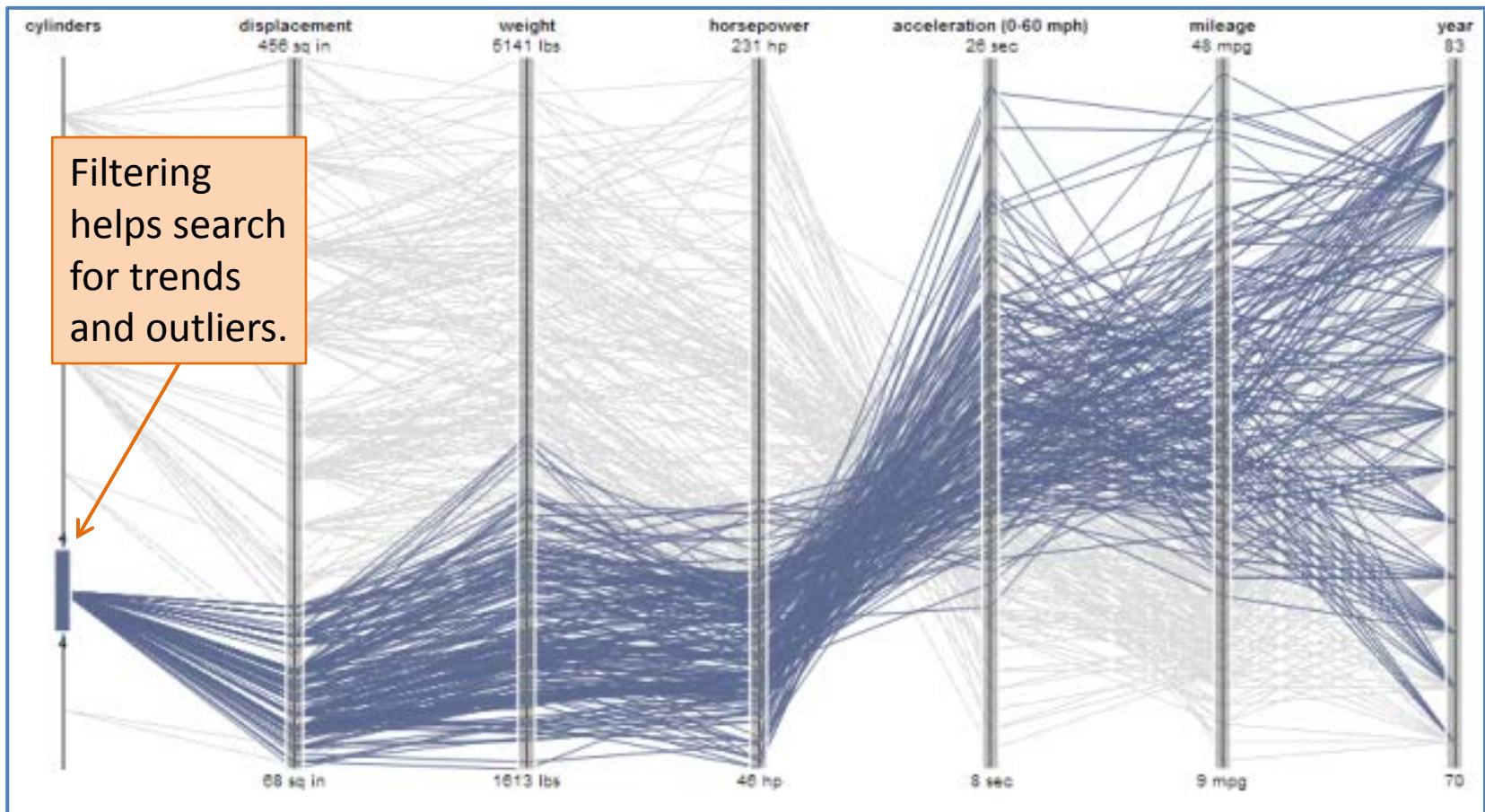
Scatter plots can produce simple maps without the need for GIS software

Visualization of multiple variables



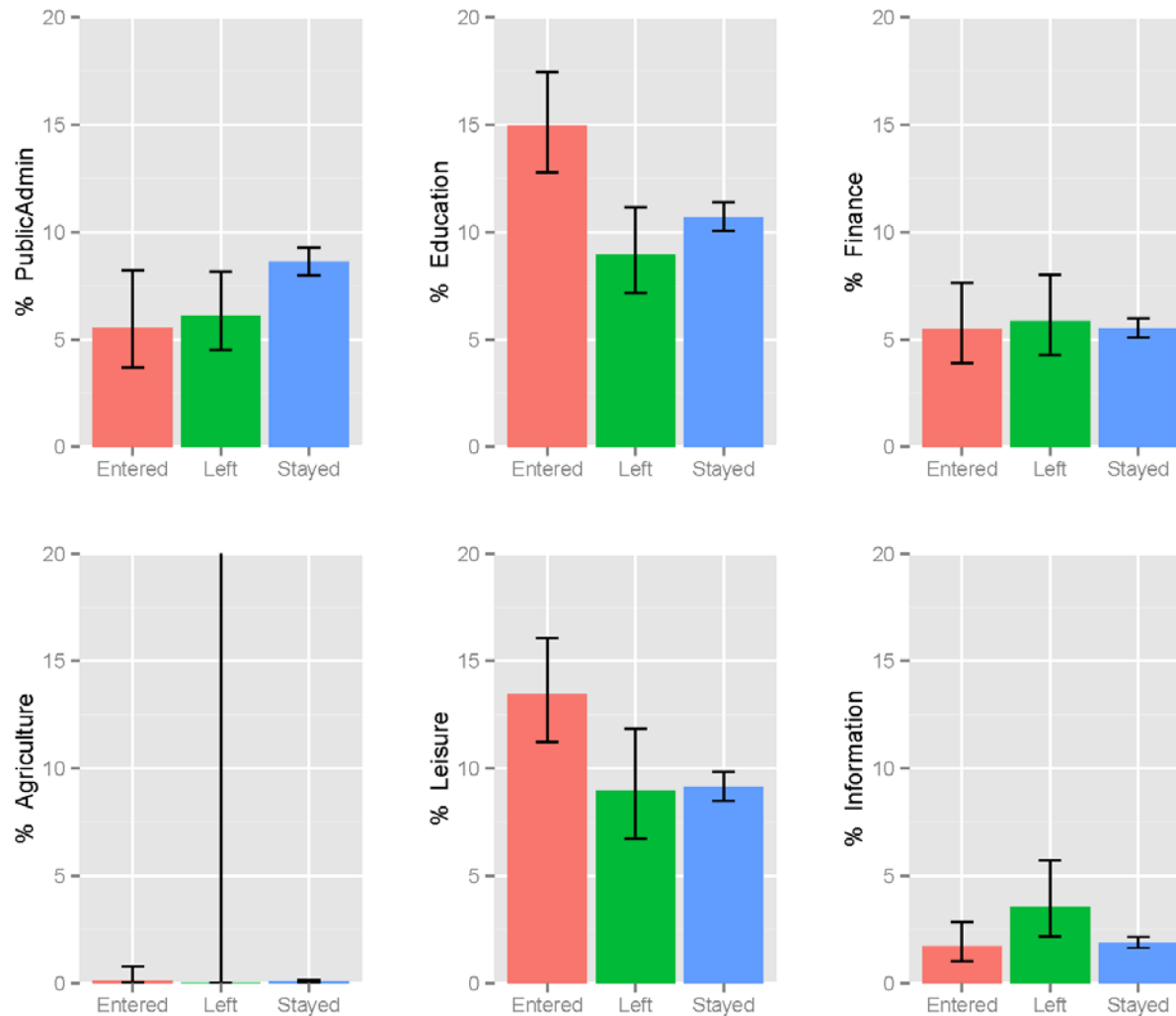
Multivariate techniques

Parallel plots or “spaghetti” plots show all the data in a single view. Each line represents a single record through all variables (columns)



Multivariate techniques

Small multiples (“facets”) make basic plotting techniques more powerful.



Employment of individuals (with 95% confidence), entering, leaving, or staying in Baltimore City 2008-2012
Draft version (excerpt). BNIA 2014. Based on U.S. Census ACS Public Use Microsample (PUMS).

The Baltimore Neighborhood Indicators Alliance (BNIA) is an organization committed to promoting, supporting, and helping people make better decisions using accurate, reliable, and accessible data and indicators to improve the quality of life in Baltimore City neighborhoods.



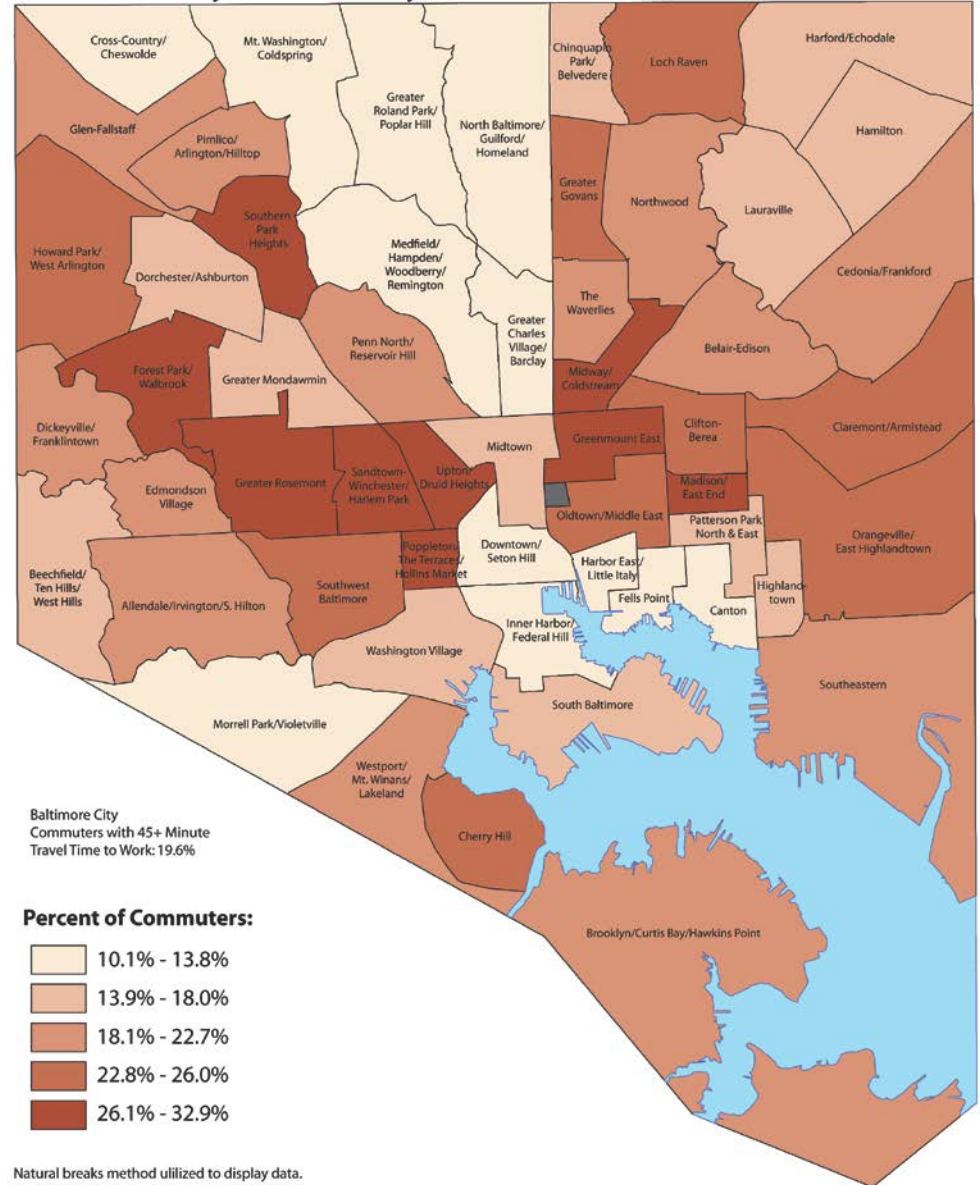
- 50+ data sources
- 150+ indicators
- Neighborhood level
- 12 years of data
- Baltimore city focus
- Specialized in mapping

<http://www.bnijfi.org/>

Baltimore City Percent of Commuters Spending 45+ Minutes Travelling to Work By Community Statistical Area, 2007-2011

Travel Time 2011	
	Avg
Growing CSAs	15.8%
Declining CSAs	23.1%

CSA = community statistical area
(neighborhood)



Natural breaks method utilized to display data.
Source: ACS, 2007-2011
Map created by BNIA-JFI, 2013

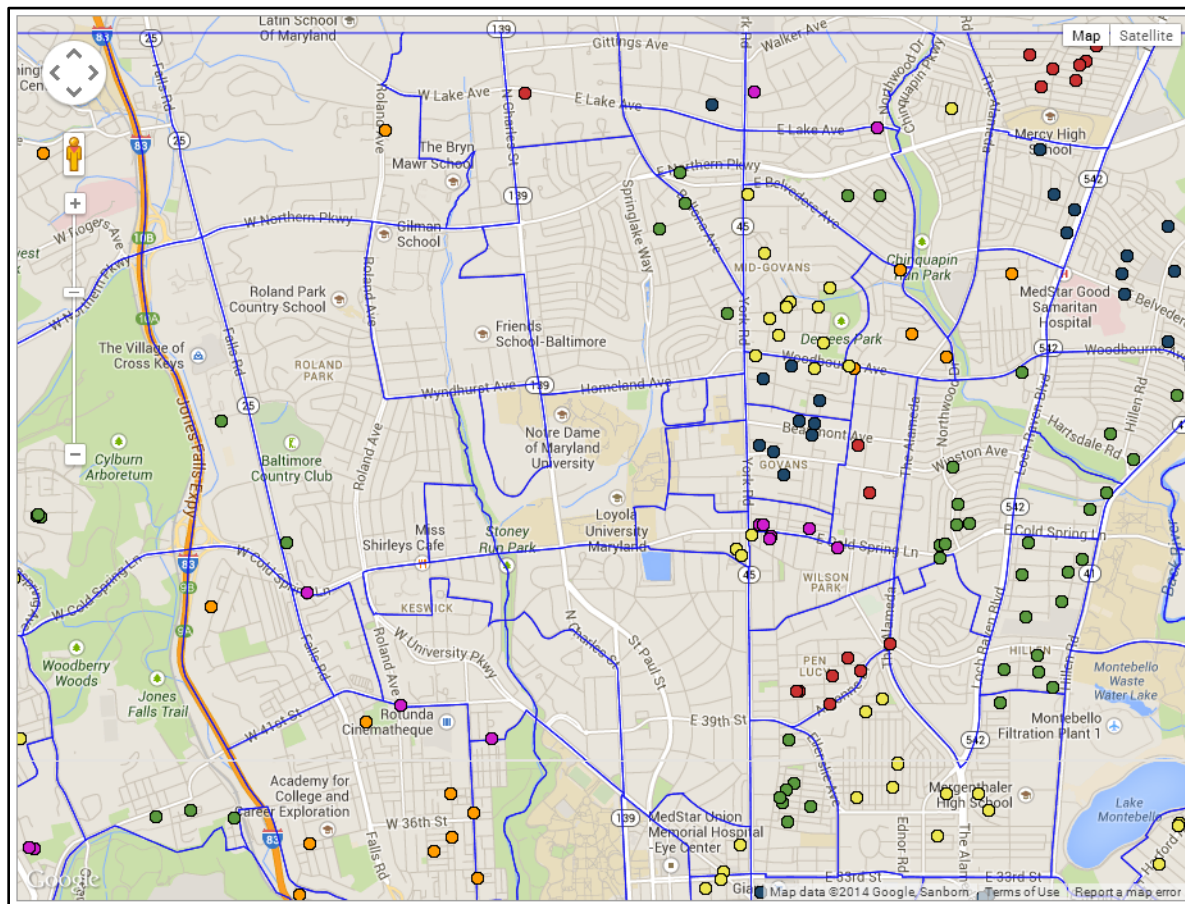
Implications for policy advocacy:

Improving public transit may support attraction & retention efforts.
Creating jobs where people live may reduce transit demand.

Foreclosures by neighborhood (Q1 2009)

“Tracking the foreclosure and home sales data provided by BNIA-JFI has enabled BHPC and our members to *better target resources and outreach efforts* around programs and other assistance, and to measure our progress.”

--Baltimore Housing Preservation Coalition, *Vital Signs 11*, page 37.

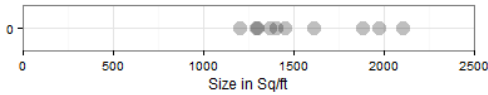


Clustered points of the same color represent foreclosures in the same neighborhood.

<http://foreclosures.bnijfi.org/>

Thank you!

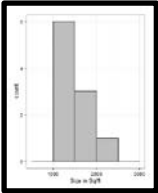
Using the ggplot2 package in R with the table named "data"



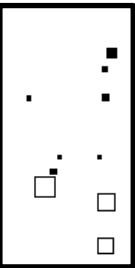
```
ggplot(data, aes(Size,0))  
+ geom_point(alpha=0.25, size=5)  
+ coord_cartesian(ylim=c(0,0),xlim=c(0,2500))  
+ theme_bw() + ylab("") + xlab("Size in Sq/ft")
```



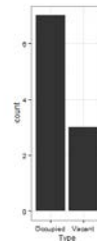
```
ggplot(data, aes(Size,0))  
+ geom_point(alpha=0.25, size=5, aes(color=Type))  
+ coord_cartesian(ylim=c(0,0),xlim=c(0,2500))  
+ theme_bw() + ylab("") + xlab("Size in Sq/ft")
```



```
ggplot(data, aes(Size))  
+ geom_histogram(binwidth=500,fill='gray', colour='black')  
+ theme_bw() + xlab("Size in Sq/ft")
```



```
ggplot(data, aes(x=X, y=Y, shape=Type, size=Size))  
+ geom_point() + scale_shape_manual(values=c(15,0))  
+ coord_cartesian(ylim=c(0,1.5),xlim=c(0,.8))  
+ theme_classic() + scale_size(range = c(1, 6))
```



```
ggplot(data,aes(Type))  
+ geom_bar()  
+ theme_bw()
```