

# MIS 0855 Fall 2016 – Data Science

## *Week 5 – Data Visualization (I)*

**Min-Seok Pang**

Management Information Systems  
Fox School of Business, Temple University  
minspang@temple.edu  
*Sep. 29<sup>th</sup>, 2016*

# Data Visualization

- You have got a good data source.
- You have several hypotheses and sensible theories to support them.
- You have got an interesting analysis result.
- How to present our data and analysis outcome?



<http://golancourses.net/2010spring/01/18/looking-outwards-data-visualization/>

# Peak Break-Up Times

According to Facebook status updates



David McCandless & Lee Byron  
InformationIsBeautiful.net / LeeByron.com

source: searches for "we broke up because"  
taken from the infographic ultrabook  
The Visual Miscellaneum

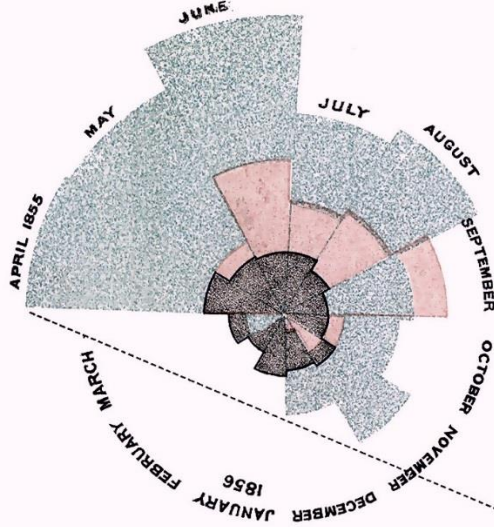
# Data Visualization Principles

- Principle #1 – Have a clear message to deliver.
- Principle #2 – Select a delivery method based on the message to deliver.
- Principle #3 – Use as little ink as possible.
- Principle #4 – Do not mislead your audience.

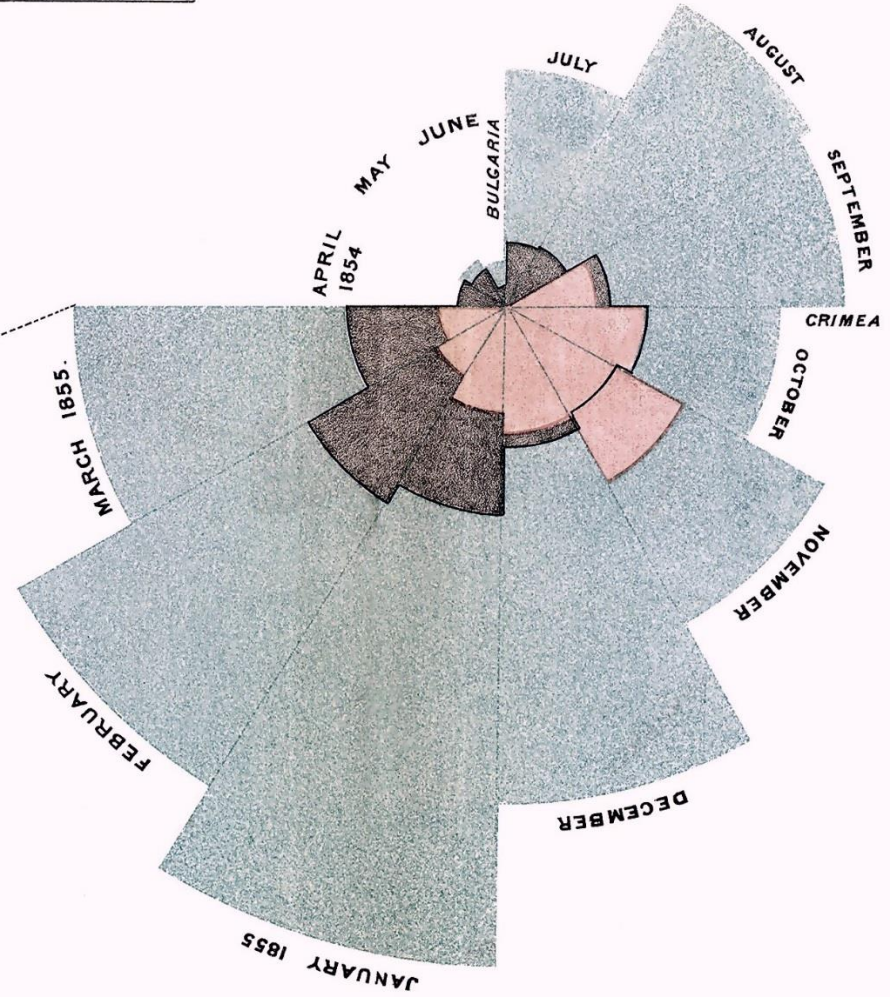
# What messages does this tell?

## 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 of the circle representing the month.

The blue wedges measured from the centre of the circle represent area of the preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges the deaths from other causes.

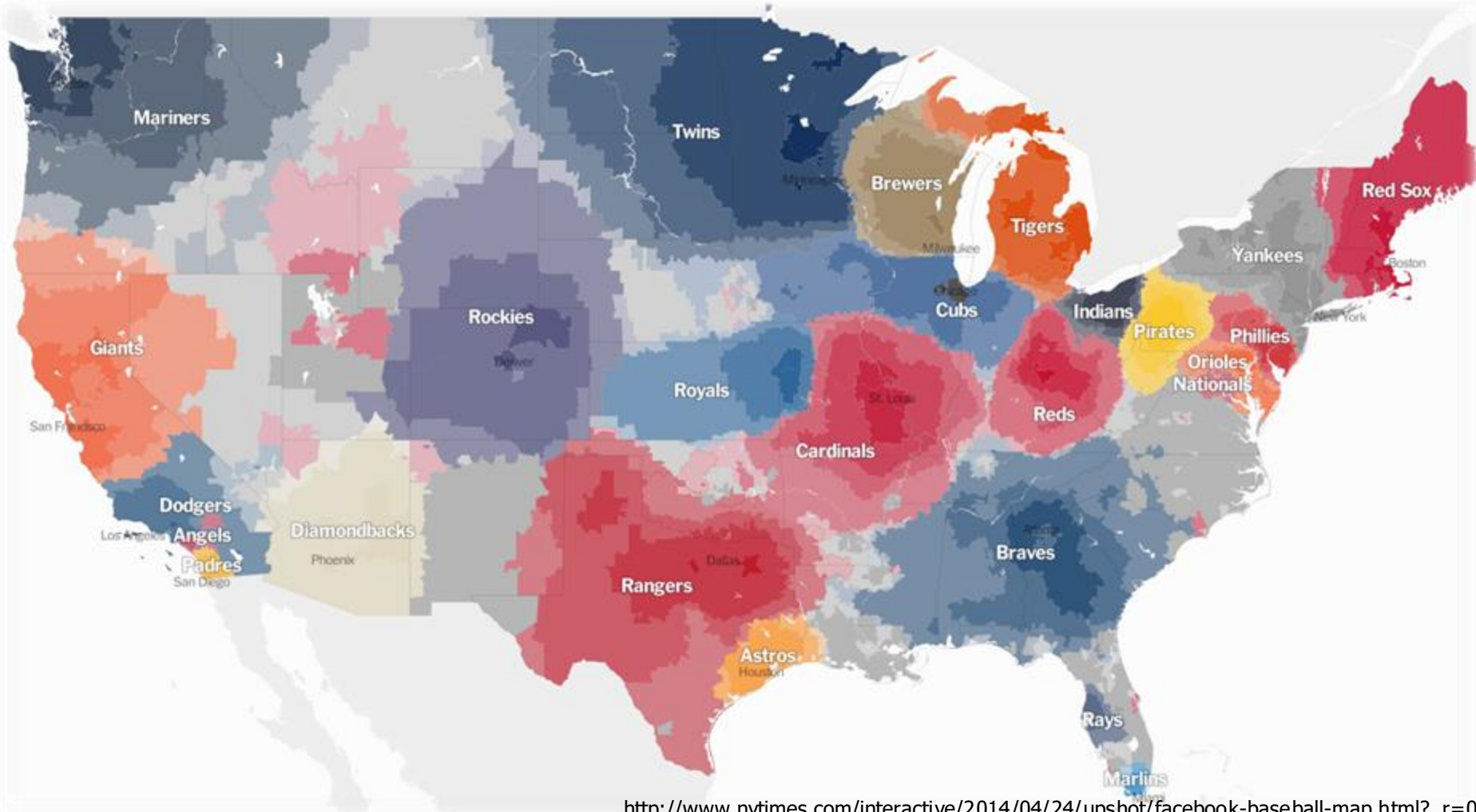
**Blue** – preventable diseases  
**Red** – wounds  
**Black** – other

The black line across the red triangle in Nov. 1854 marks the boundary of the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855 the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

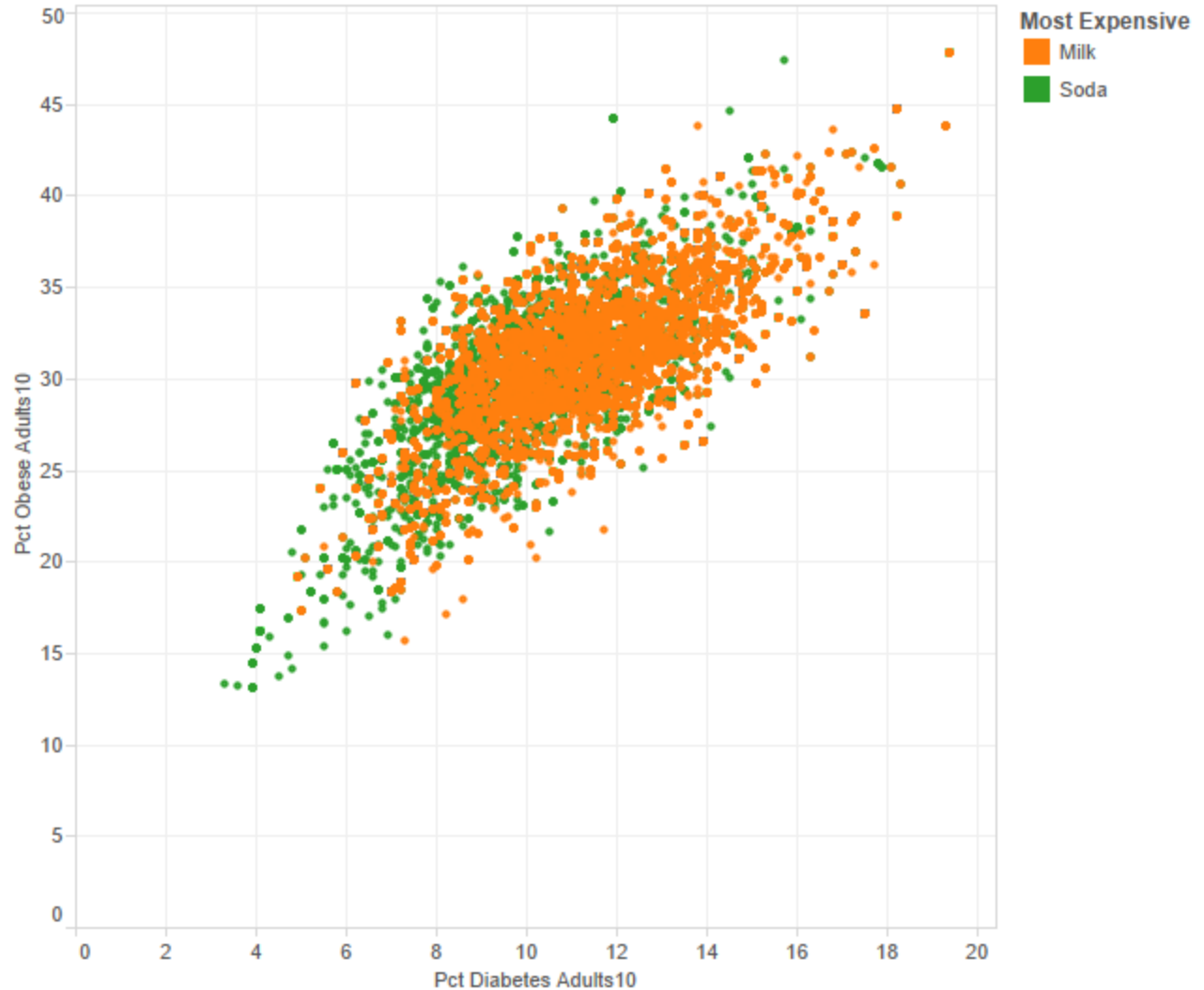
# What messages does this tell?



[http://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html?\\_r=0](http://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html?_r=0)

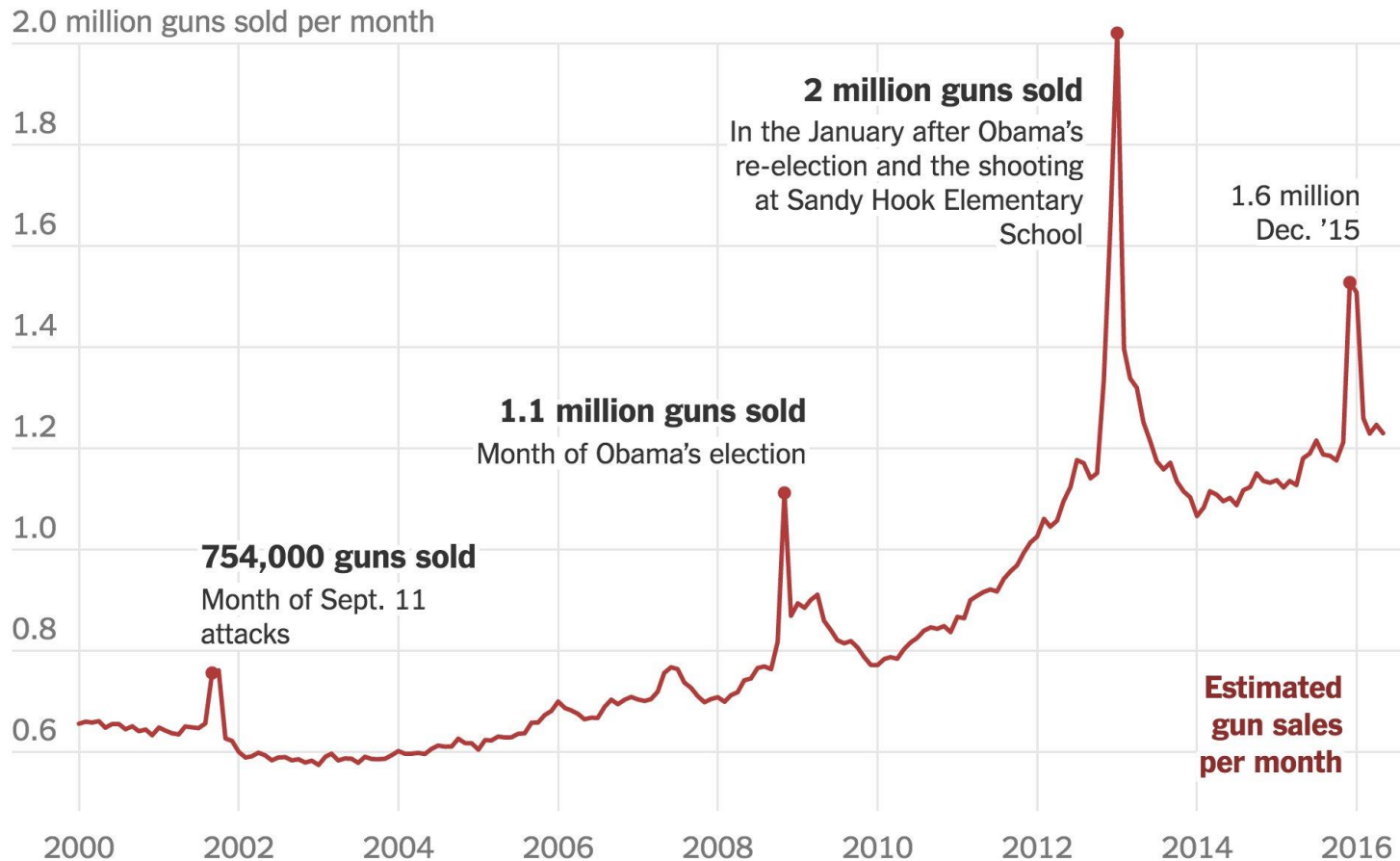
**What  
messages  
does  
this tell?**

Impact of Beverage Prices



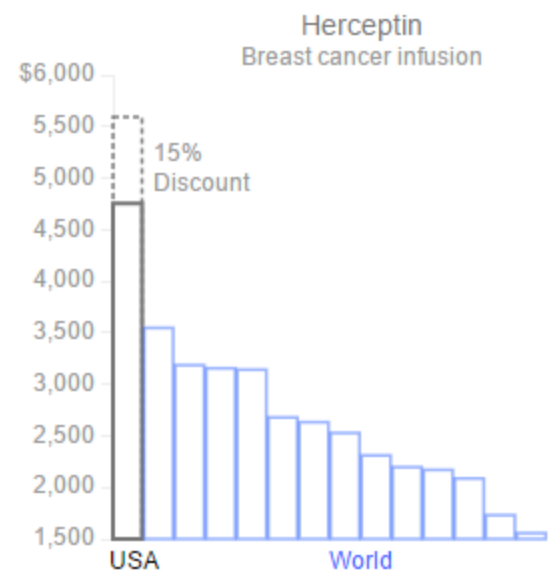
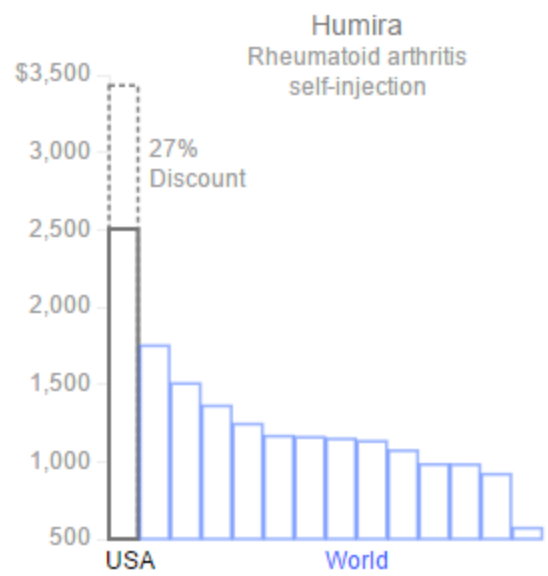
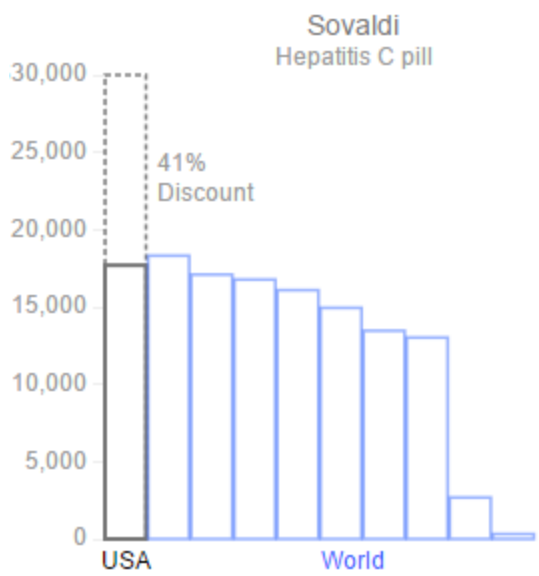
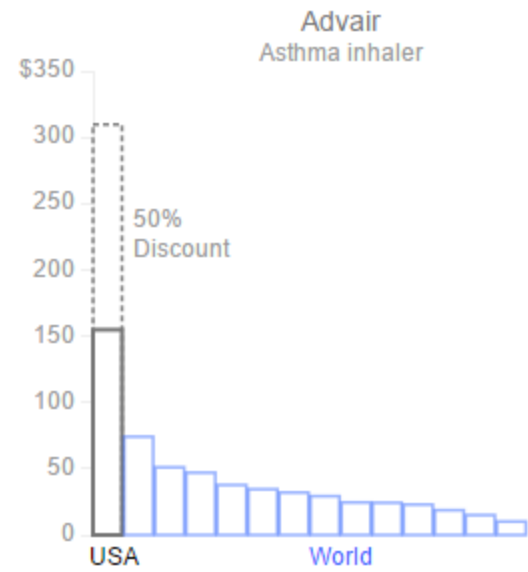
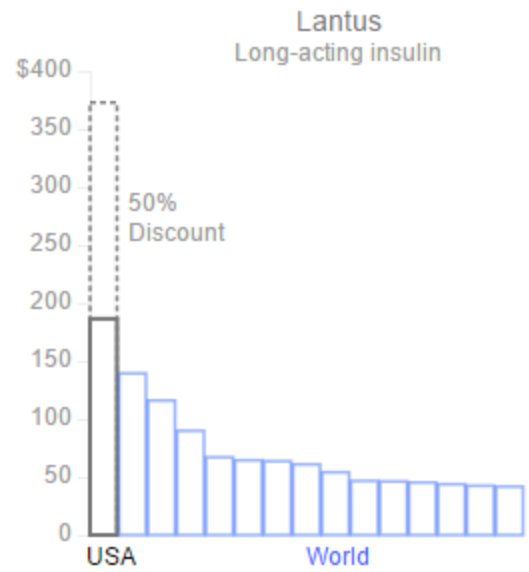
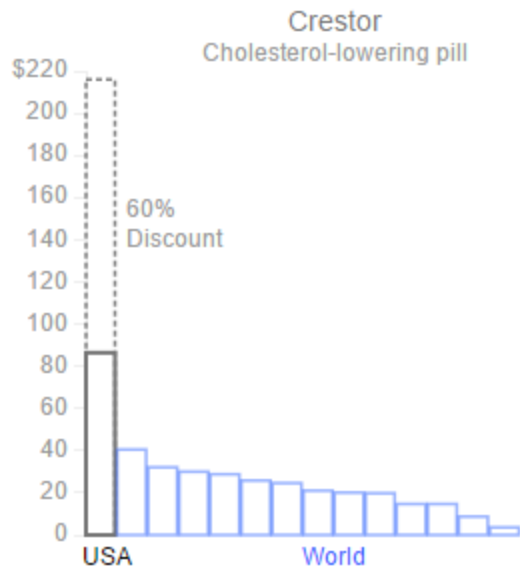
Pct Diabetes Adults10 vs. Pct Obese Adults10. Color shows details about Most Expensive. The view is filtered on Most Expensive, which excludes Null.

# What messages does this tell?



[http://www.nytimes.com/interactive/2015/12/10/us/gun-sales-terrorism-obama-restrictions.html?smid=pl-share&\\_r=0](http://www.nytimes.com/interactive/2015/12/10/us/gun-sales-terrorism-obama-restrictions.html?smid=pl-share&_r=0)

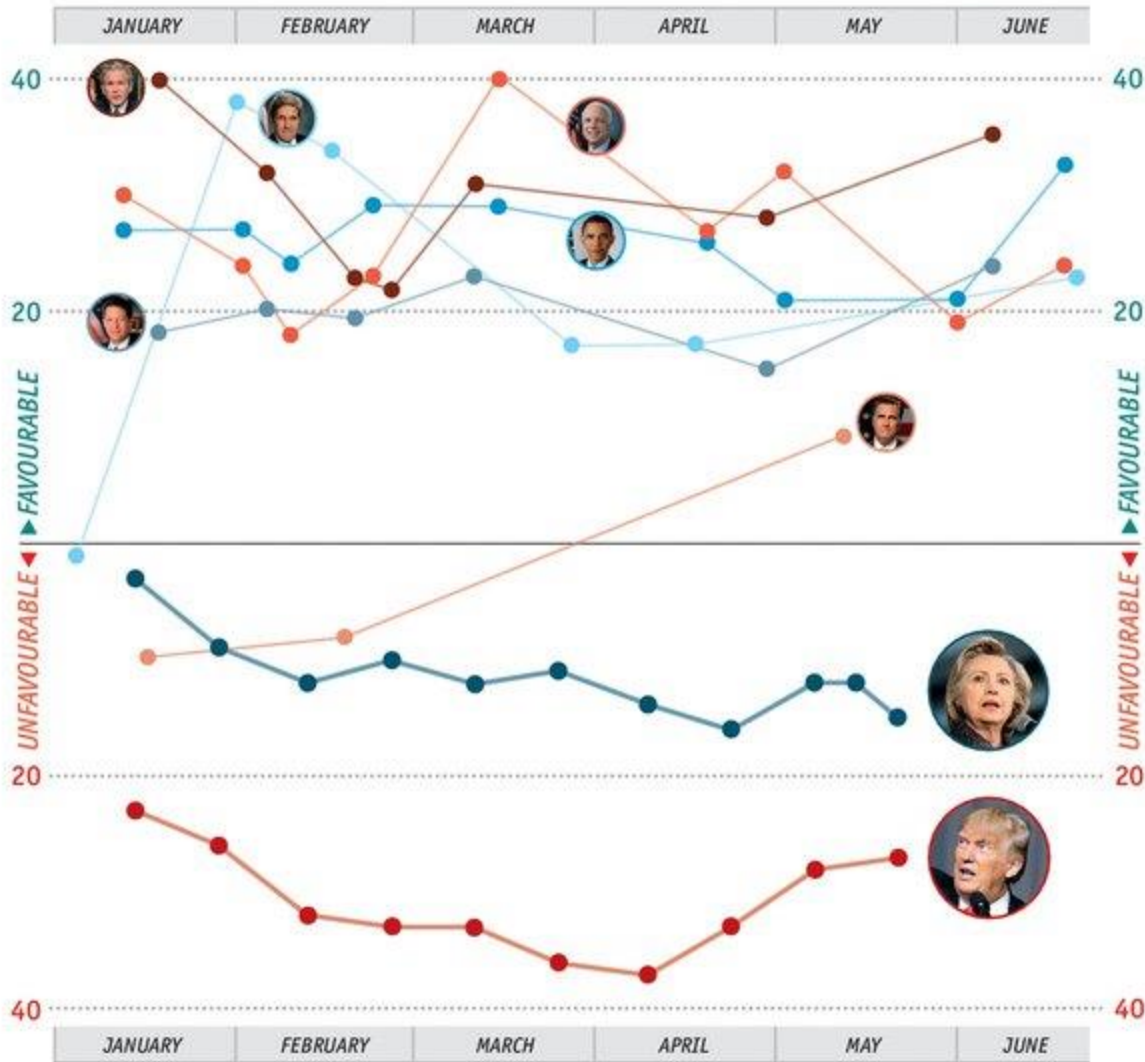




# Fear and loathing

Net favourability ratings\*, election years 2000-16, %

- 2000 ● George W. Bush
- 2004 ● John Kerry
- 2008 ● Barack Obama
- 2012 ● Mitt Romney
- 2016 ● Hillary Clinton
- Al Gore
- John McCain
- Donald Trump



Source: Gallup Credit: AP

\*Percentage point difference between favourable minus unfavourable responses

<http://www.scoopnest.com/user/ECONdailycharts/735485061693362177>

Economist.com

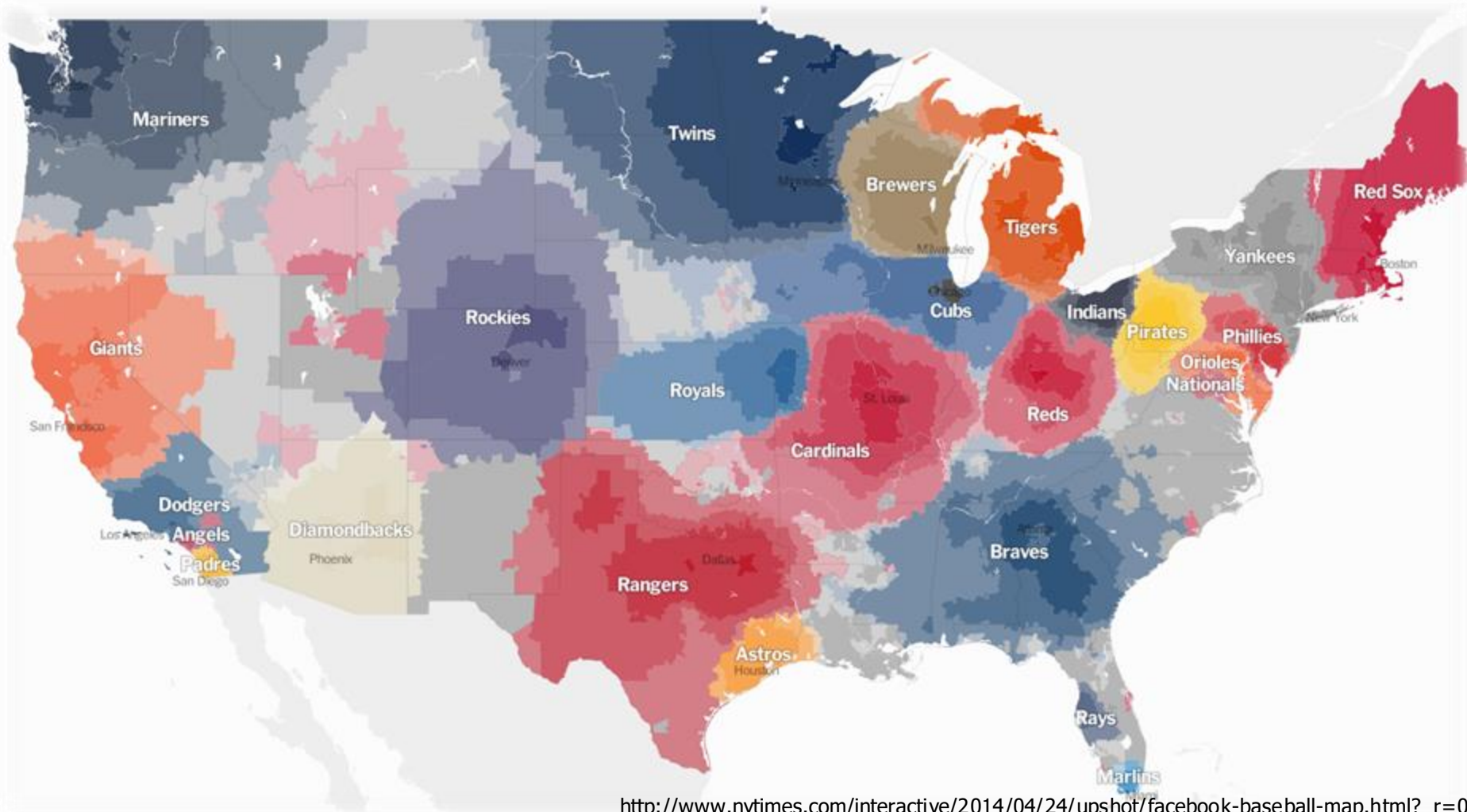
## Principle #2

- Select a delivery method based on the message to deliver.
- Table vs. Graph?
- Bar vs. Line vs. Scatter vs. Pie?

# Why a Table?

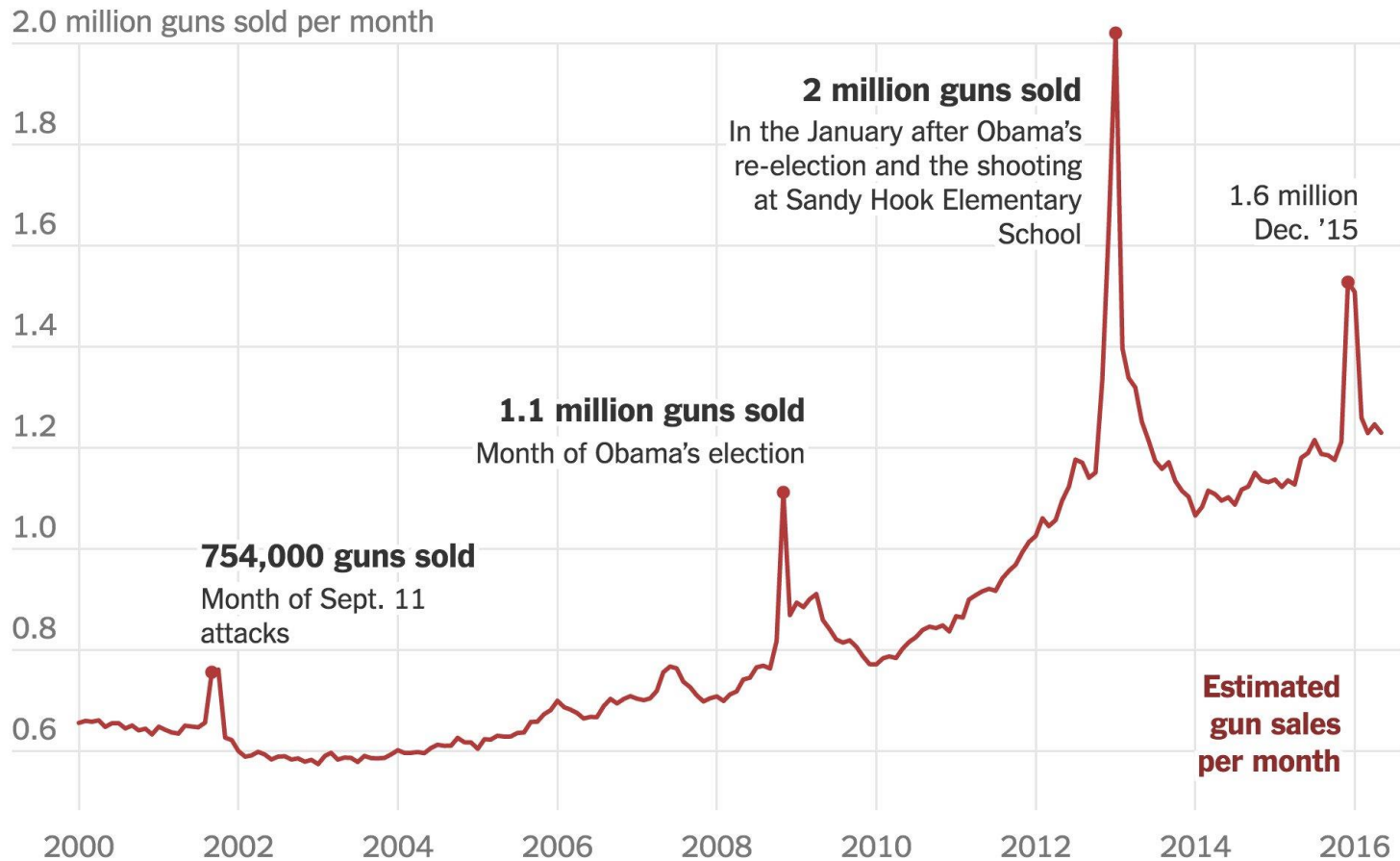
Candidates	Campaign Feb. 1 - Feb. 29			Super-PACs Feb. 1 - Feb. 29			Total 2015 - Feb. 29, 2016 Raised
	Raised	Spent	Cash on Hand	Raised	Spent	Cash on Hand	
Bernie Sanders	<b>\$43.5m</b>	<b>\$40.9m</b>	<b>\$17.2m</b>	<b>\$0.0m</b>	<b>\$0.0m</b>	<b>\$0.0m</b>	<b>\$139.8m</b>
Hillary Clinton	<b>29.5</b>	<b>31.6</b>	<b>30.8</b>	<b>4.9*</b>	<b>5.3*</b>	<b>44.5*</b>	<b>229.5*</b>
Ted Cruz	<b>11.9</b>	<b>17.5</b>	<b>8.0</b>	<b>5.9</b>	<b>11.5</b>	<b>20.0</b>	<b>120.5</b>
Marco Rubio	<b>9.6</b>	<b>9.2</b>	<b>5.5</b>	<b>25.0</b>	<b>27.2</b>	<b>3.4</b>	<b>112.0</b>
Donald Trump	<b>9.2</b>	<b>9.5</b>	<b>1.3</b>	<b>0.0</b>	<b>0.0</b>	<b>0.0</b>	<b>34.7</b>
Ben Carson	<b>5.7</b>	<b>5.0</b>	<b>4.9</b>	<b>*</b>	<b>*</b>	<b>*</b>	<b>77.3*</b>
John Kasich	<b>3.4</b>	<b>3.6</b>	<b>1.3</b>	<b>3.4</b>	<b>3.2</b>	<b>2.5</b>	<b>22.4</b>
Jeb Bush	<b>1.2</b>	<b>3.6</b>	<b>0.5</b>	<b>0.2</b>	<b>8.5</b>	<b>16.1</b>	<b>153.6</b>
Chris Christie	<b>0.4</b>	<b>0.9</b>	<b>0.3</b>	<b>0.5</b>	<b>1.1</b>	<b>0.1</b>	<b>28.7</b>
Carly Fiorina	<b>0.2</b>	<b>1.8</b>	<b>1.8</b>	<b>0.03</b>	<b>1.2</b>	<b>1.4</b>	<b>26.2</b>

# Why a Map?



[http://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html?\\_r=0](http://www.nytimes.com/interactive/2014/04/24/upshot/facebook-baseball-map.html?_r=0)

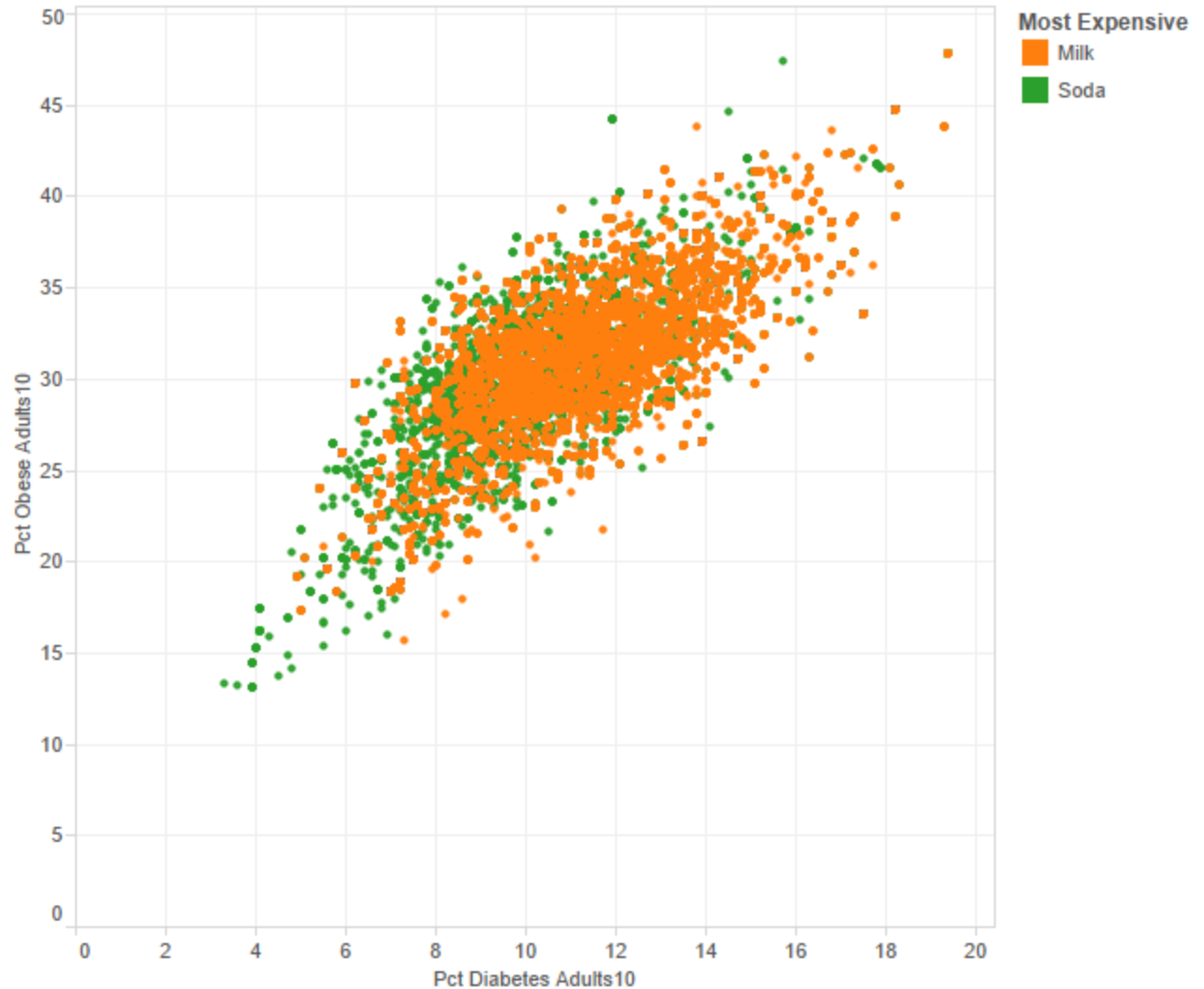
# Why a Line Chart?



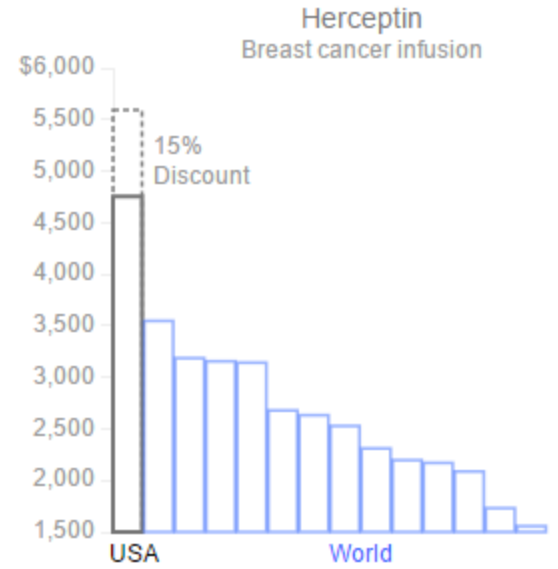
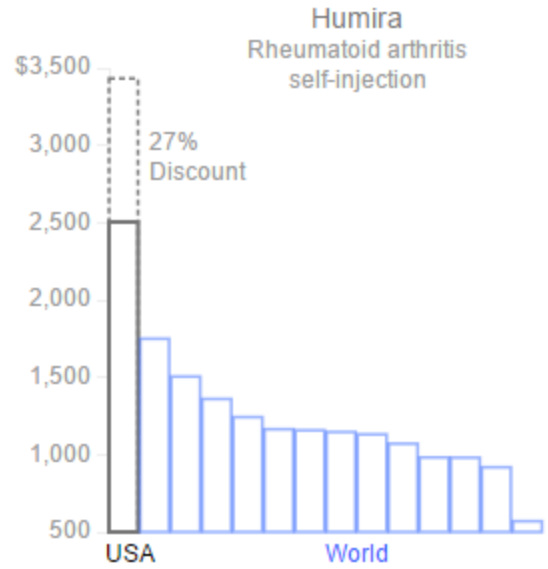
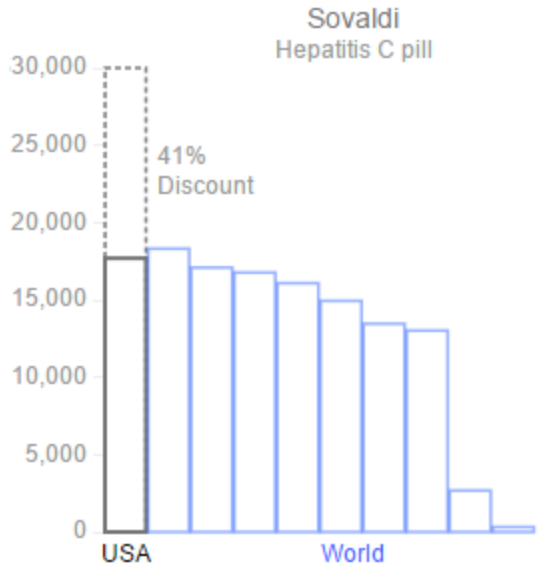
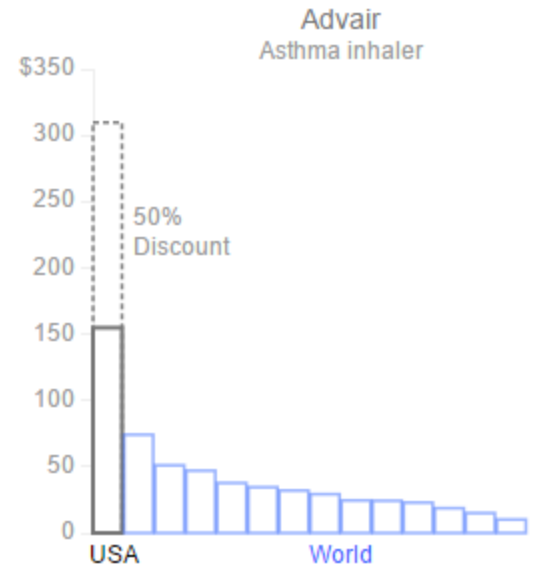
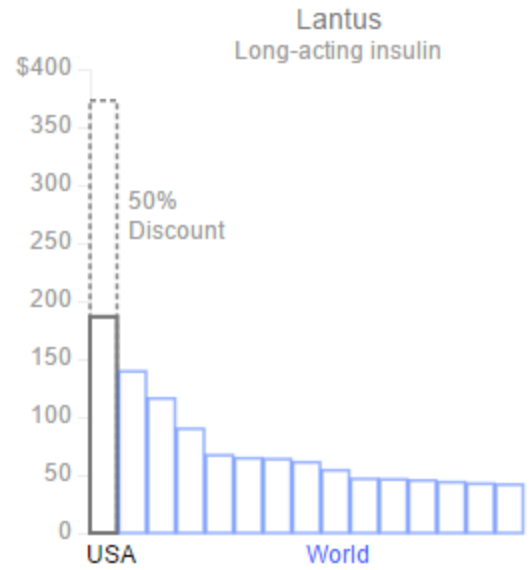
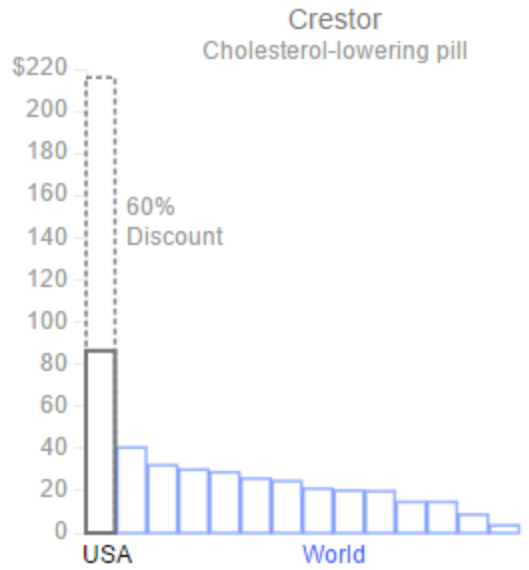
[http://www.nytimes.com/interactive/2015/12/10/us/gun-sales-terrorism-obama-restrictions.html?smid=pl-share&\\_r=0](http://www.nytimes.com/interactive/2015/12/10/us/gun-sales-terrorism-obama-restrictions.html?smid=pl-share&_r=0)

# Why a Scatterplot?

Impact of Beverage Prices



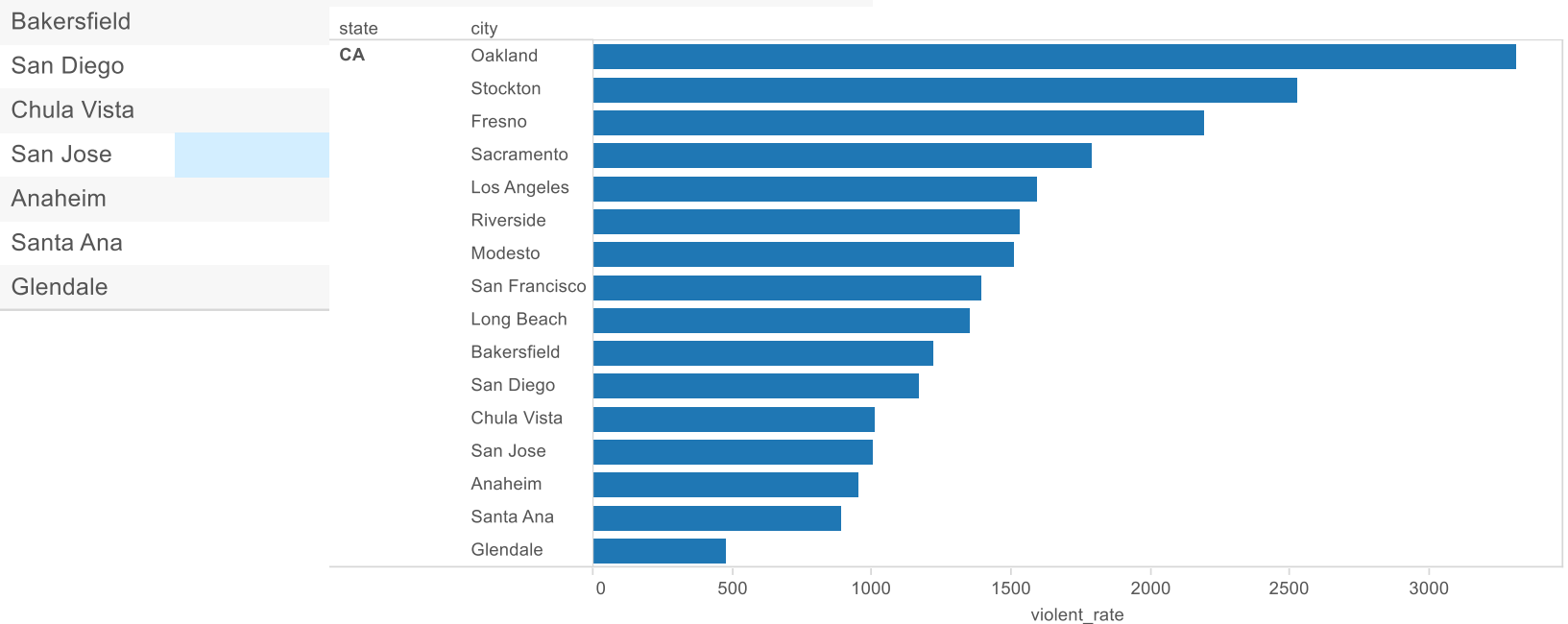
Pct Diabetes Adults10 vs. Pct Obese Adults10. Color shows details about Most Expensive. The view is filtered on Most Expensive, which excludes Null.





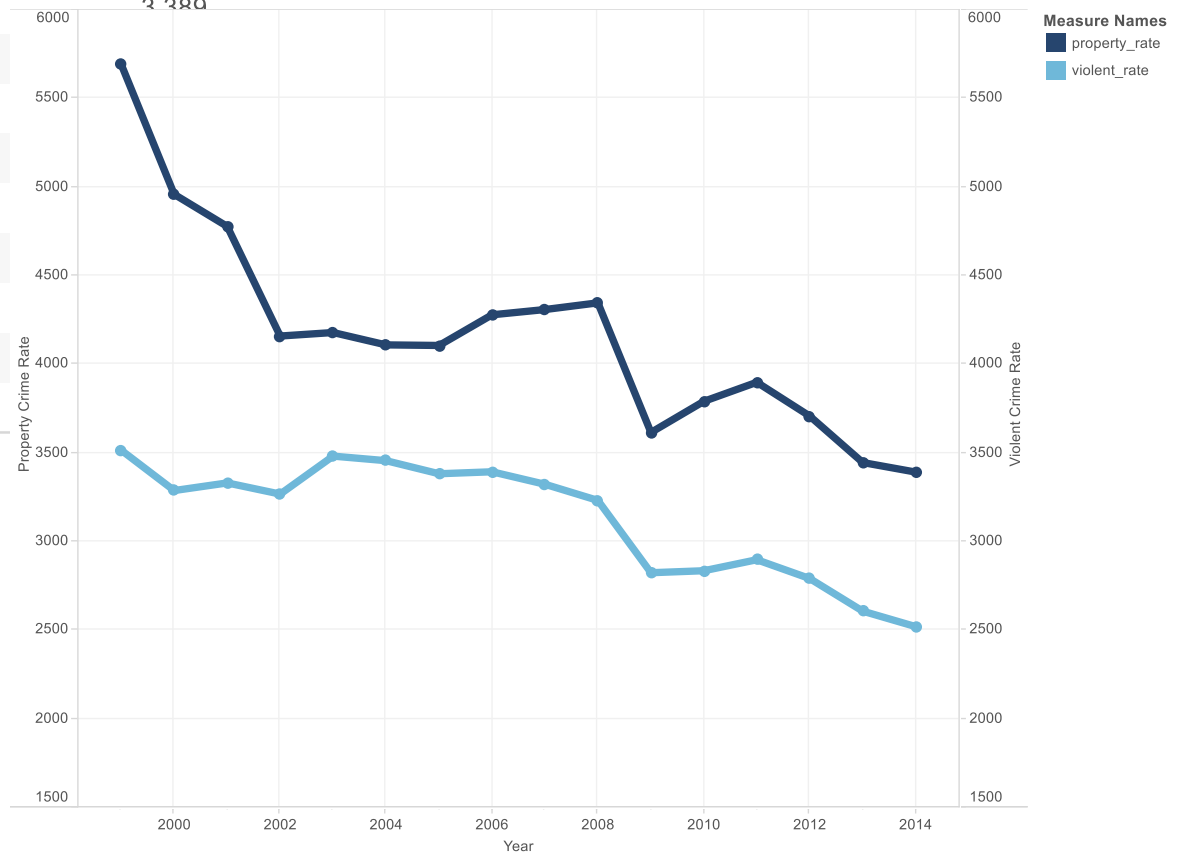
state	city	Avg. murder_rate	Avg. violent_rate	Avg. property_rate
CA	Oakland	26	3,312	5,658
	Stockton	14	2,527	5,936
	Fresno	9	2,189	5,100
	Sacramento	9	1,789	4,955
	Los Angeles	10	1,592	2,698
	Riverside	5	1,526	3,826
	Modesto	7	1,513	5,460
	San Francisco	9	1,391	4,703
	Long Beach	8	1,351	2,831
	Bakersfield			

# Which One Is More Effective?



# Which One Is More Effective?

Year of year	property_rate	violent_rate
2002	4,155	3,264
2003	4,175	3,478
2004	4,105	3,452
2005	4,102	3,378
2006	4,275	3,380
2007	4,305	
2008	4,342	
2009	3,611	
2010	3,787	
2011	3,894	
2012	3,704	
2013	3,442	
2014	3,388	



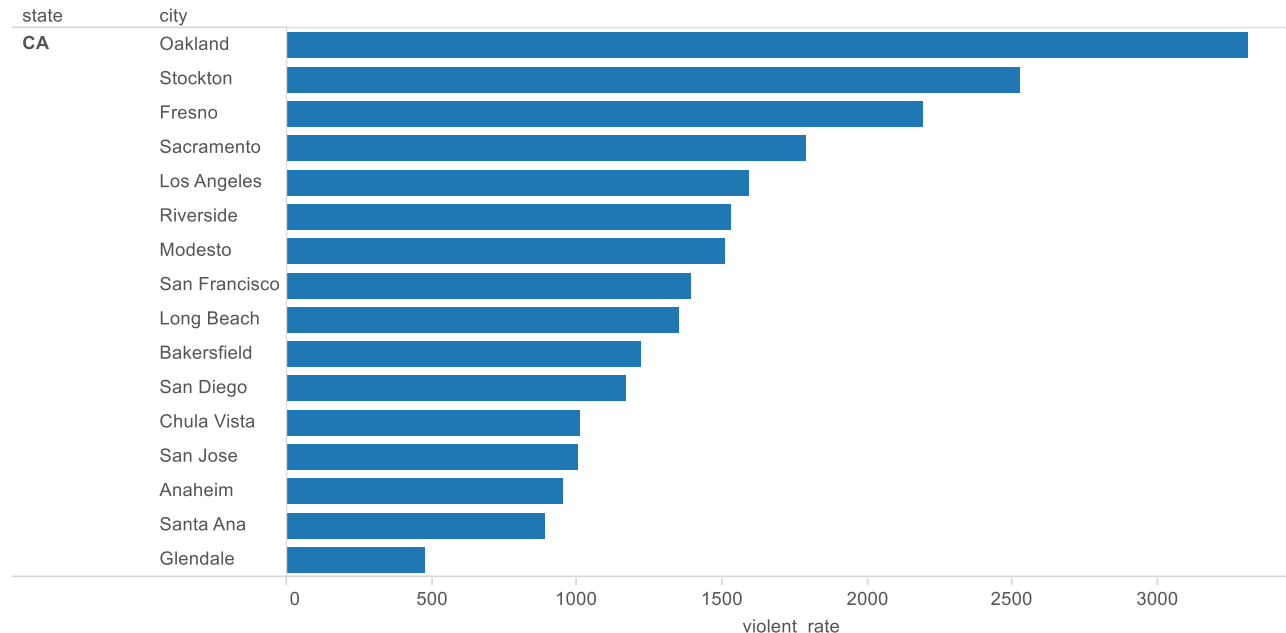
## Table is better ...

- when it needs to display individual, precise values
- when information to be communicated involves more than one unit of measure (e.g. # of people, ratio, # of days in one page)
- when summary and detail values need to be delivered

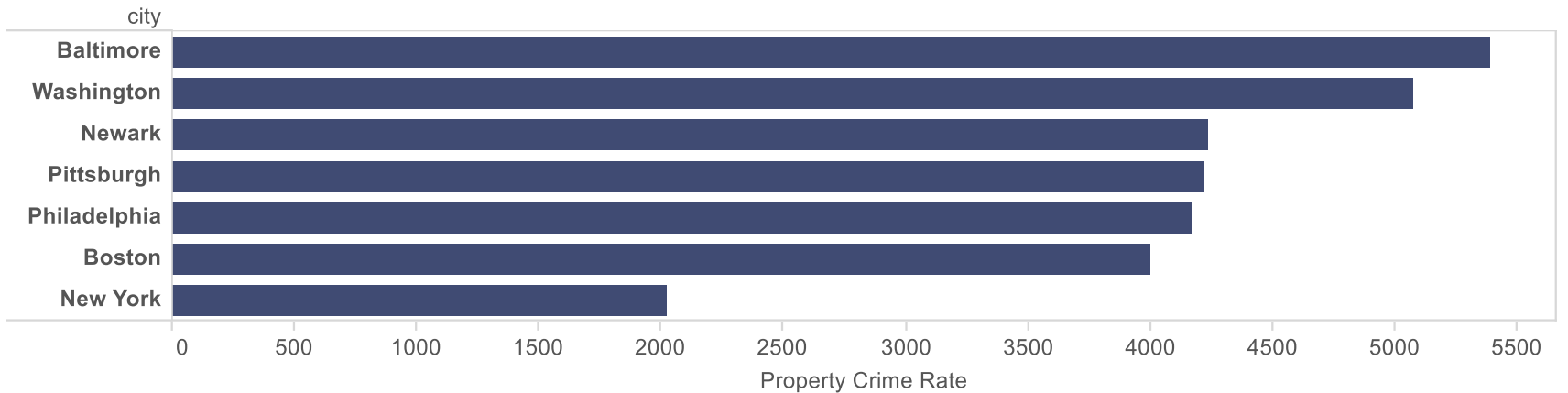
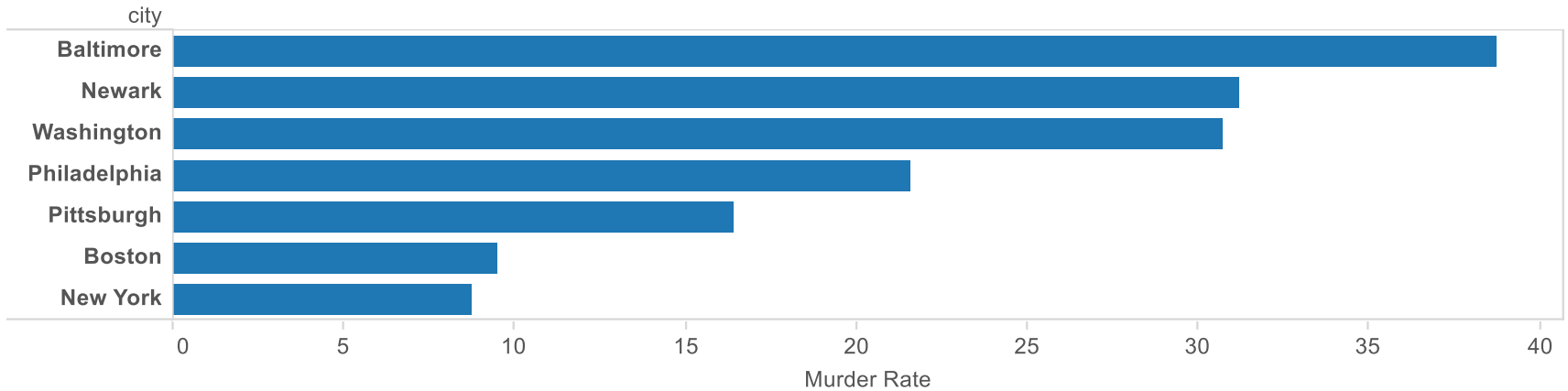
Year of year	property_rate	violent_rate
2002	4,155	3,264
2003	4,175	3,478
2004	4,105	3,452
2005	4,102	3,378
2006	4,275	3,389
2007	4,305	3,319
2008	4,342	3,228
2009	3,611	2,820
2010	3,787	2,831
2011	3,894	2,893
2012	3,704	2,787
2013	3,442	2,603
2014	3,388	2,516

## Graph is better ...

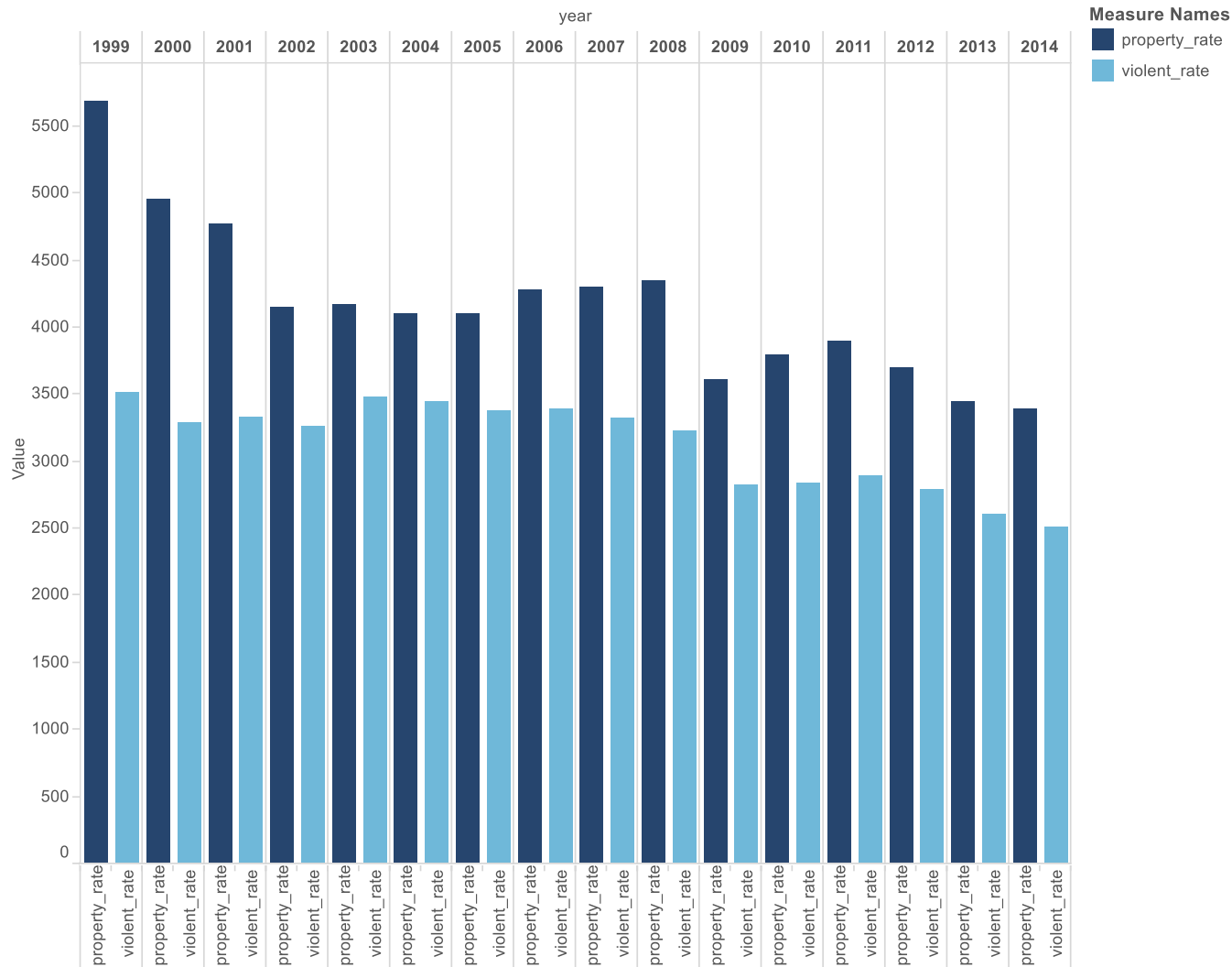
- to display a pattern or a relationship between quantitative values
- to deliver a large amount of quantitative values
- to make your audiences' life easier!



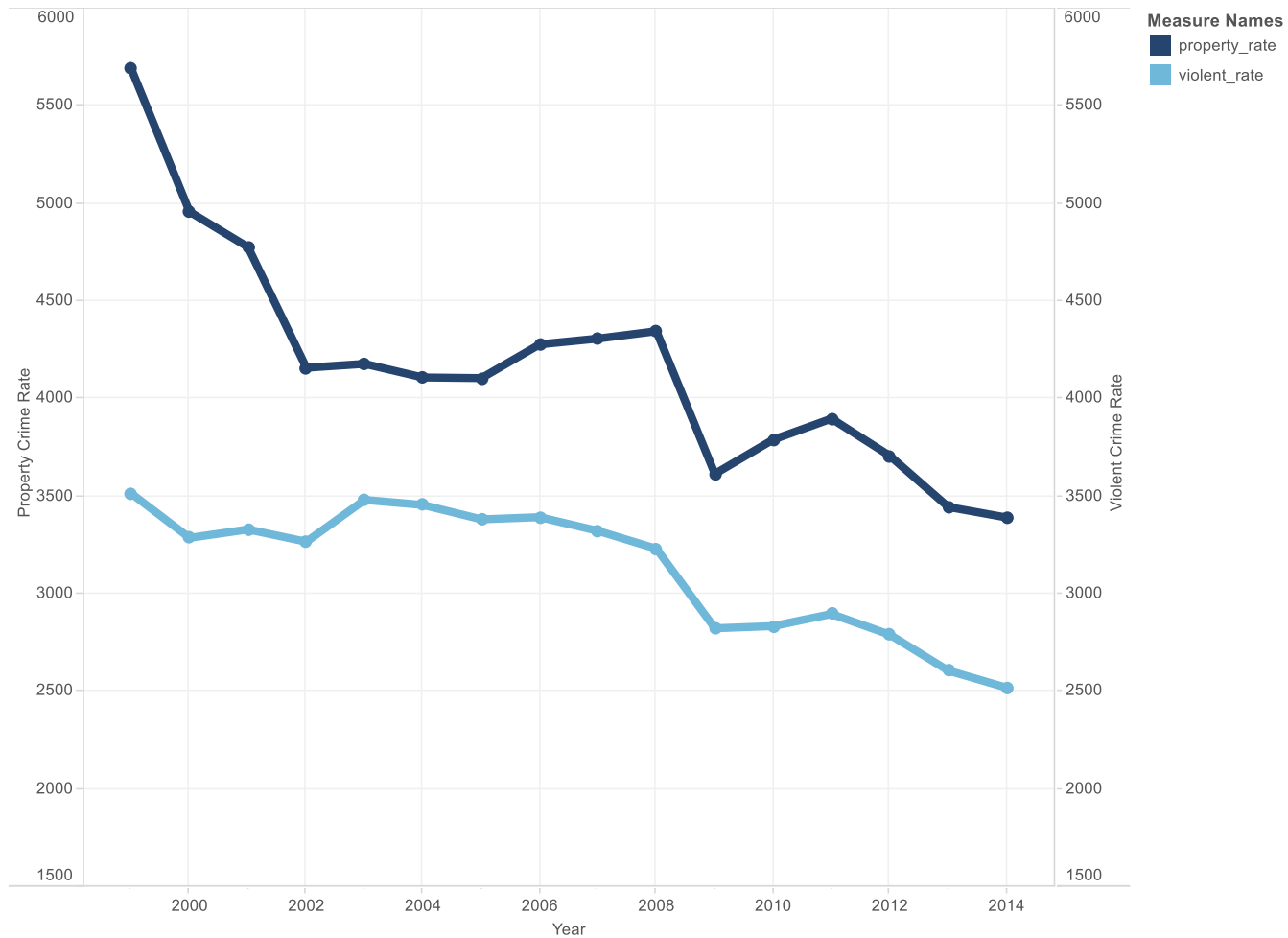
# A bar chart is good to visualize a ranking.



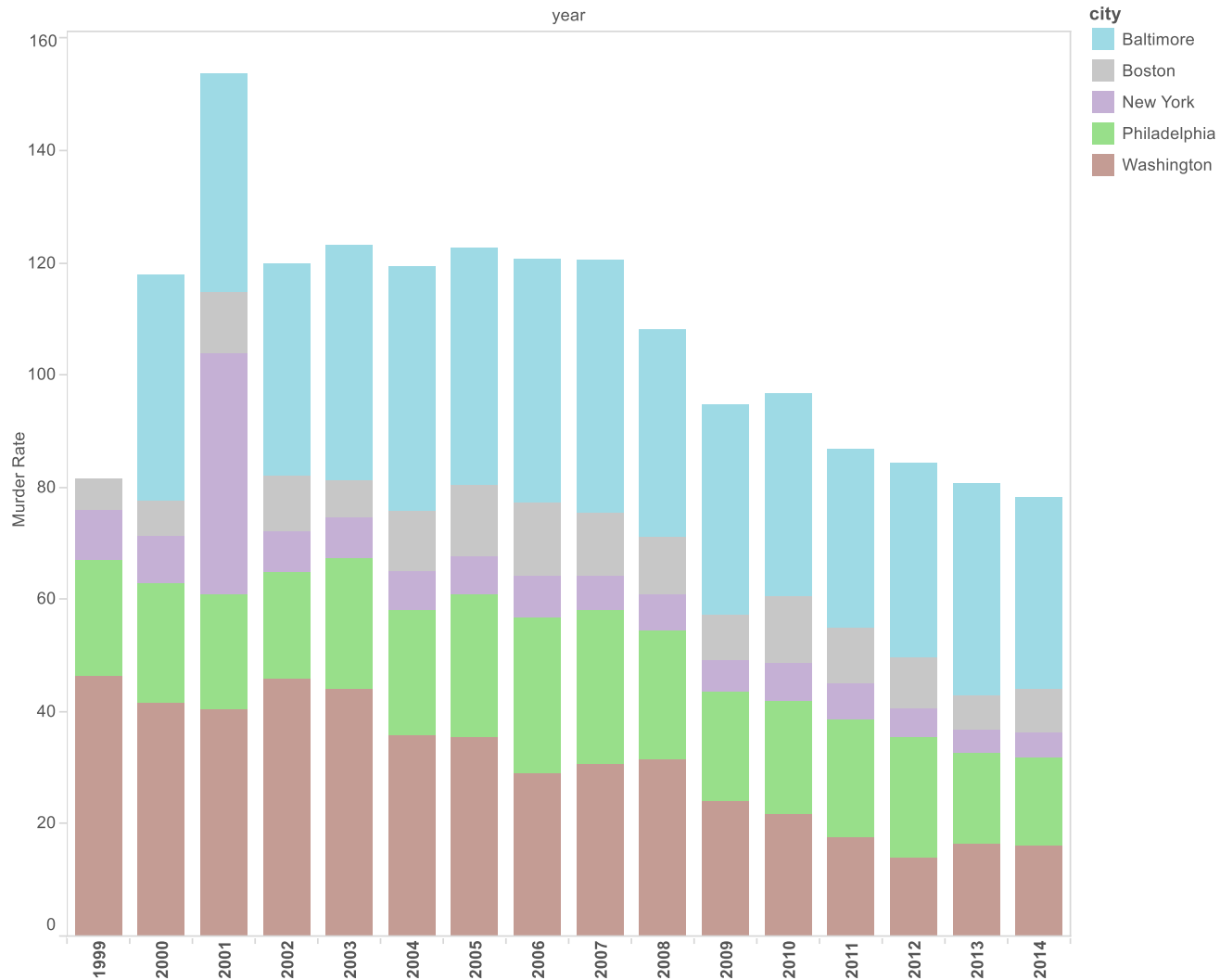
# A bar chart is NOT good to visualize a trend.



# A line chart is good to visualize a trend.

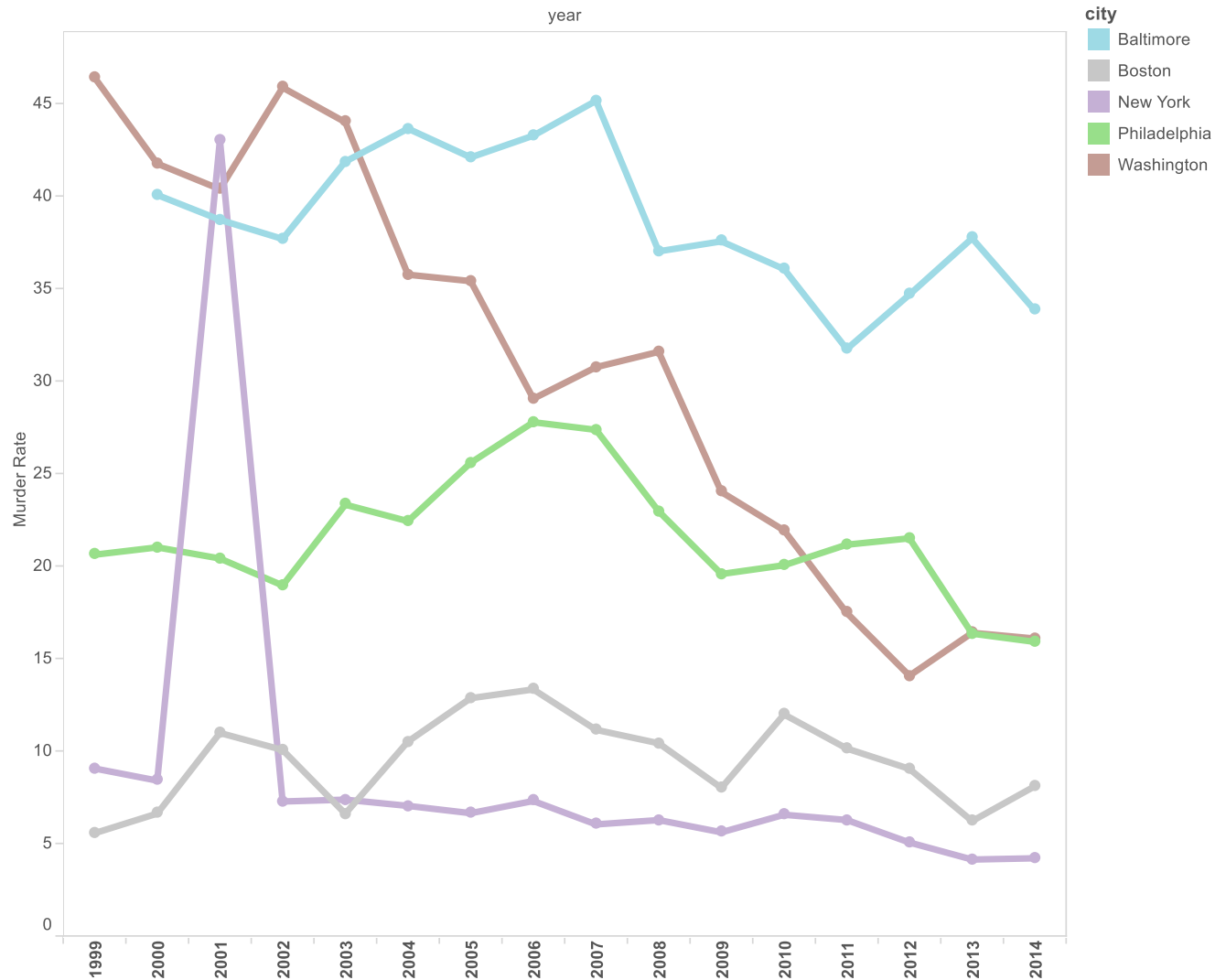


# A bar chart is NOT good to visualize a trend.

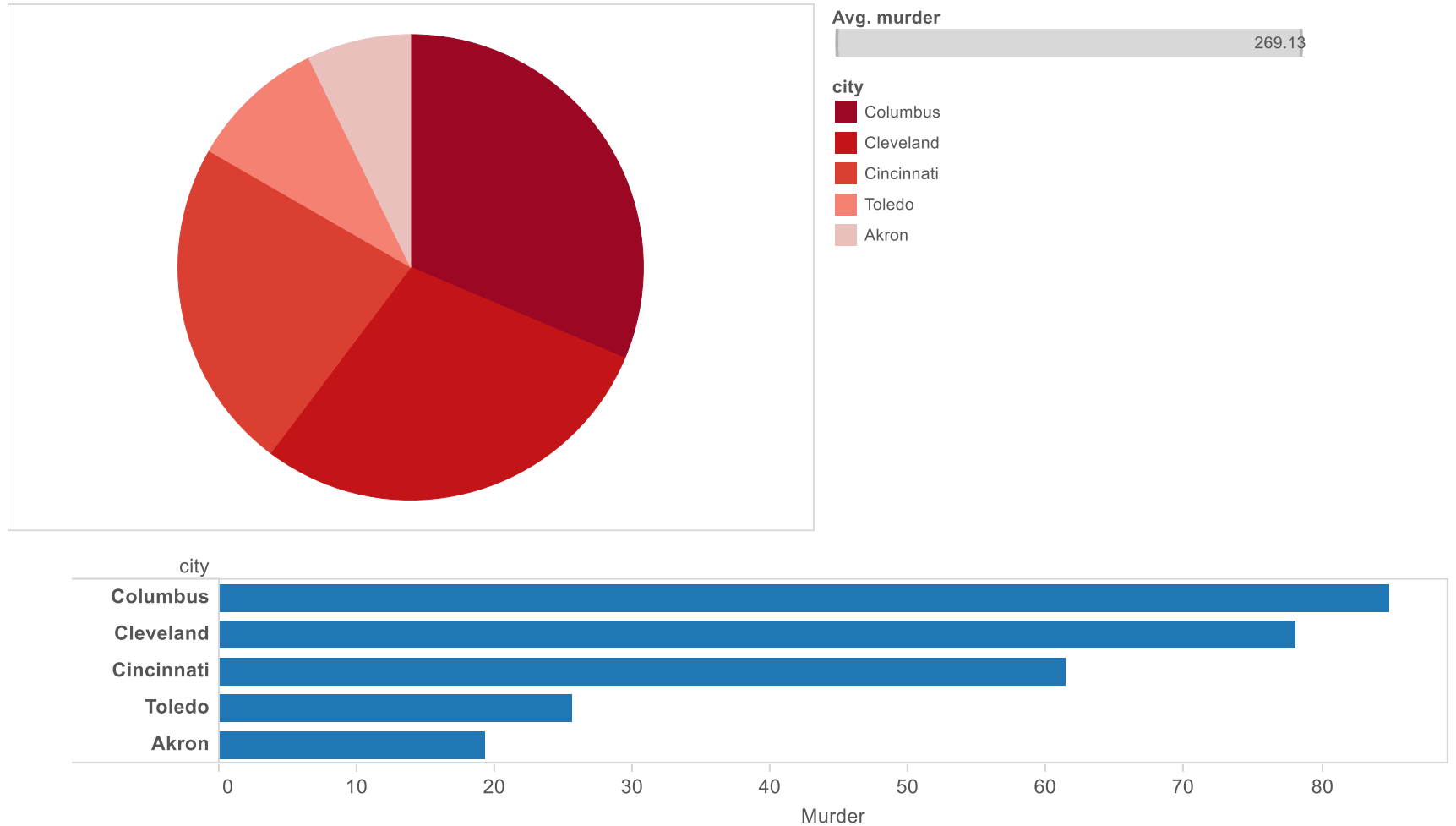




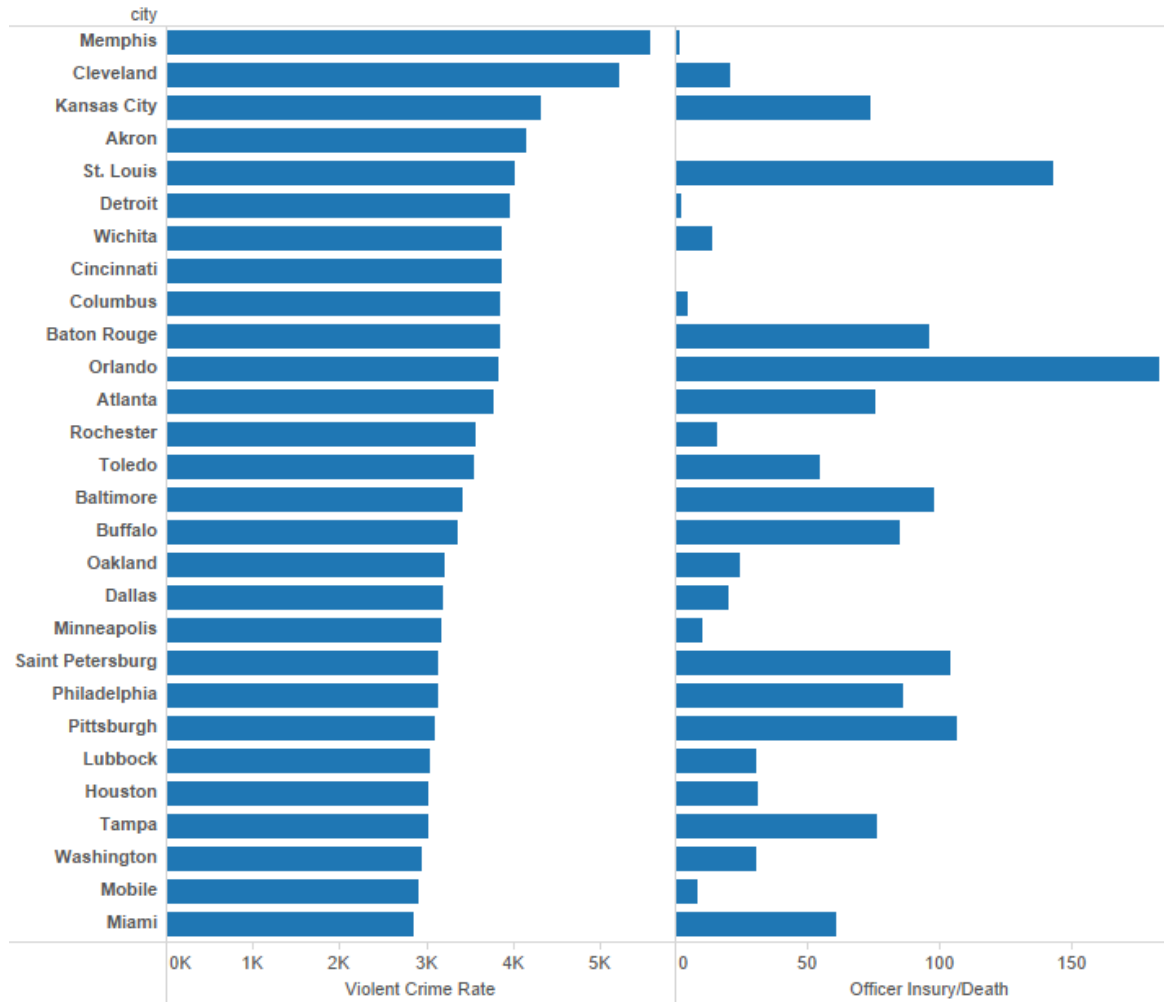
# A line chart is good to visualize a trend.



# A pie chart is good to visualize ratios.



# A bar chart is NOT good to visualize correlation.



**A scatterplot is good to visualize correlation.**

