



# DATA VISUALIZATION



# UNIT - I



# Contents

1

Introduction

2

The Importance of Context

3

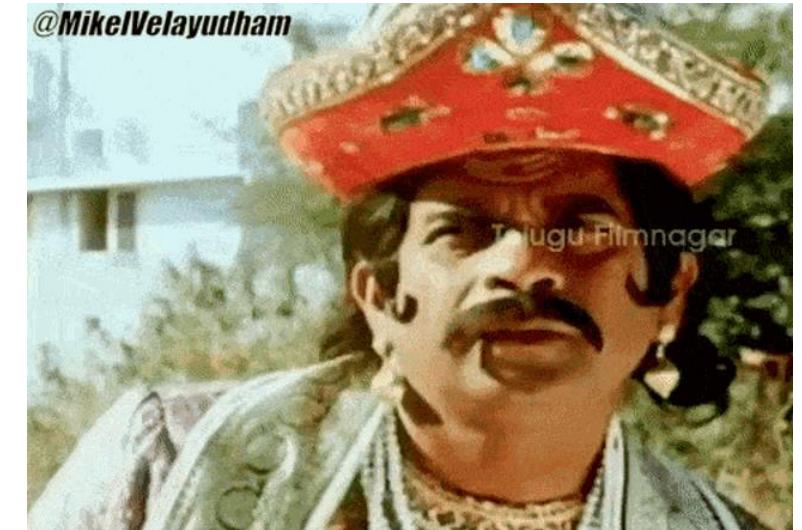
Choosing the Effective Visual



# Introduction

## Example:

- In the IPL 2023, In the Chennai Super Kings vs Lucknow Super Giants match the Chennai team player Ruturaj Gaikwad scores 57 runs in 31 balls, Convey scores 47 runs in 29 balls and in Lucknow team the player KL Rahul scored 20 runs in 18 balls, The other player Mayers scored 53 runs in just 22 balls. The result of the match was Chennai super kings won by 12 runs.





# Introduction

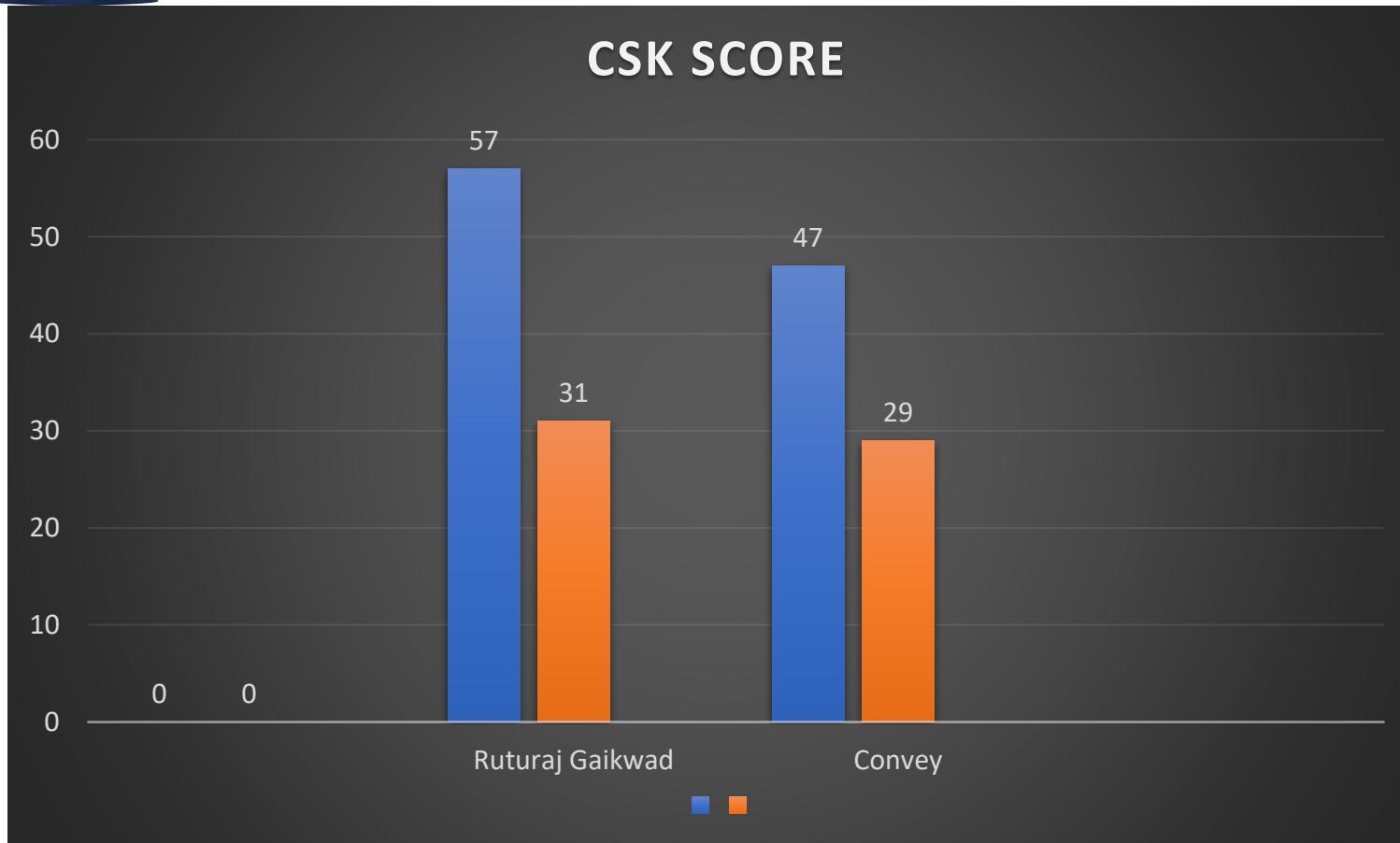
## Example:

- In the IPL 2023, In the Chennai Super Kings vs Lucknow Super Giants match the Chennai team player Ruturaj Gaikwad scores 57 runs in 31 balls, Convey scores 47 runs in 29 balls and in Lucknow team the player KL Rahul scored 20 runs in 18 balls, The other player Mayers scored 53 runs in just 22 balls. The result of the match was Chennai super kings won by 12 runs.





# Introduction





# Introduction

- Data visualization is a powerful and essential tool in the field of data analysis and communication.
- It involves the representation of data through visual elements such as charts, graphs, maps, and other visual aids to help people understand complex information, patterns, and trends.

## Applications:

- Exploration and Discovery
- Communication
- Decision-Making
- Storytelling



# Introduction

## Storytelling

- Storytelling in the context of data visualization refers to the practice of using data and visual elements to convey a narrative or message that engages and informs the audience.





# Introduction

- It's about turning raw data into a compelling story that makes complex information more relatable, understandable, and memorable.
- Effective data storytelling goes beyond just presenting numbers and facts; it aims to create an emotional connection and help the audience grasp the significance of the data.

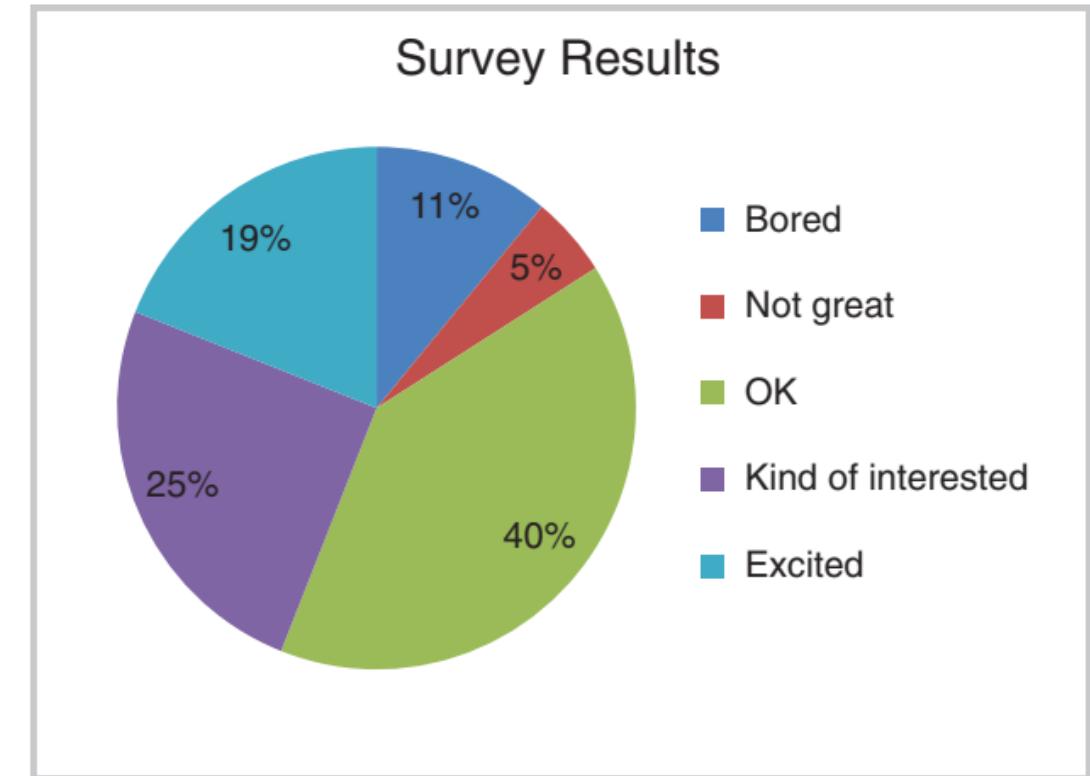




# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.

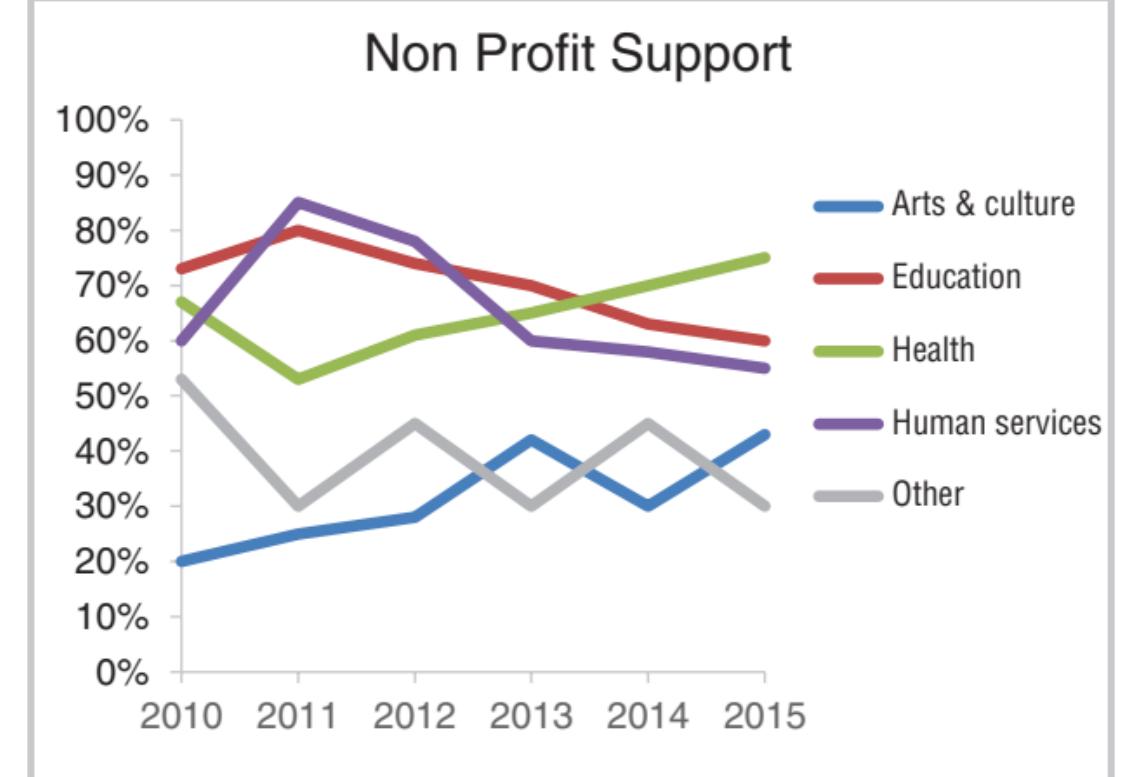




# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.

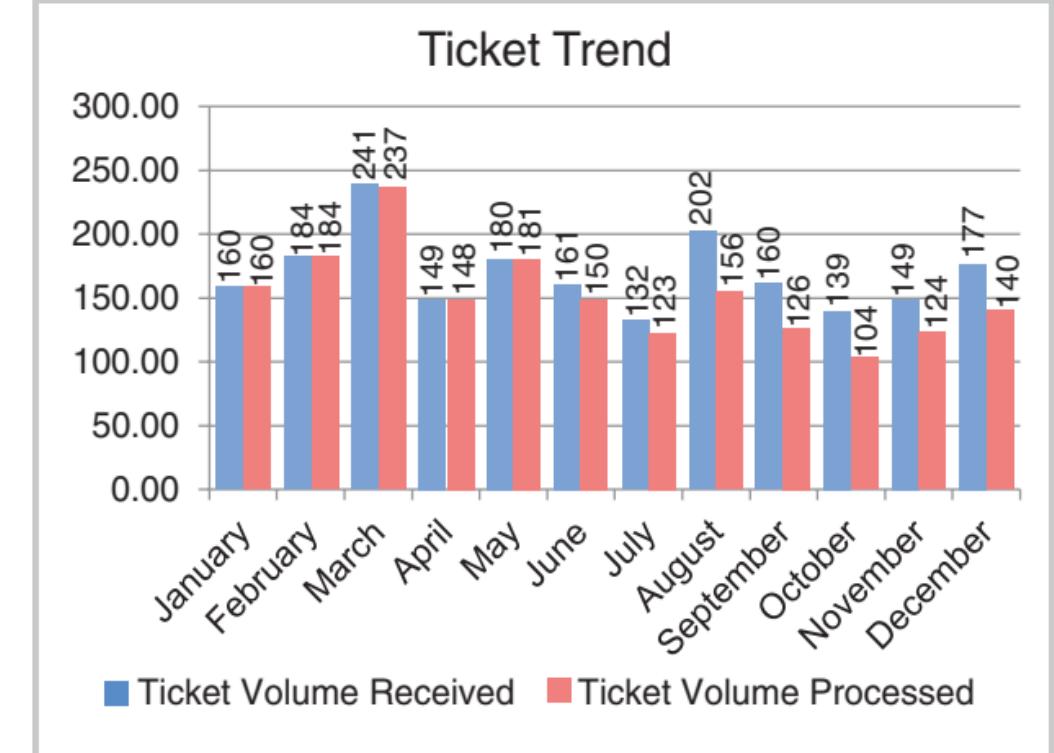




# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.



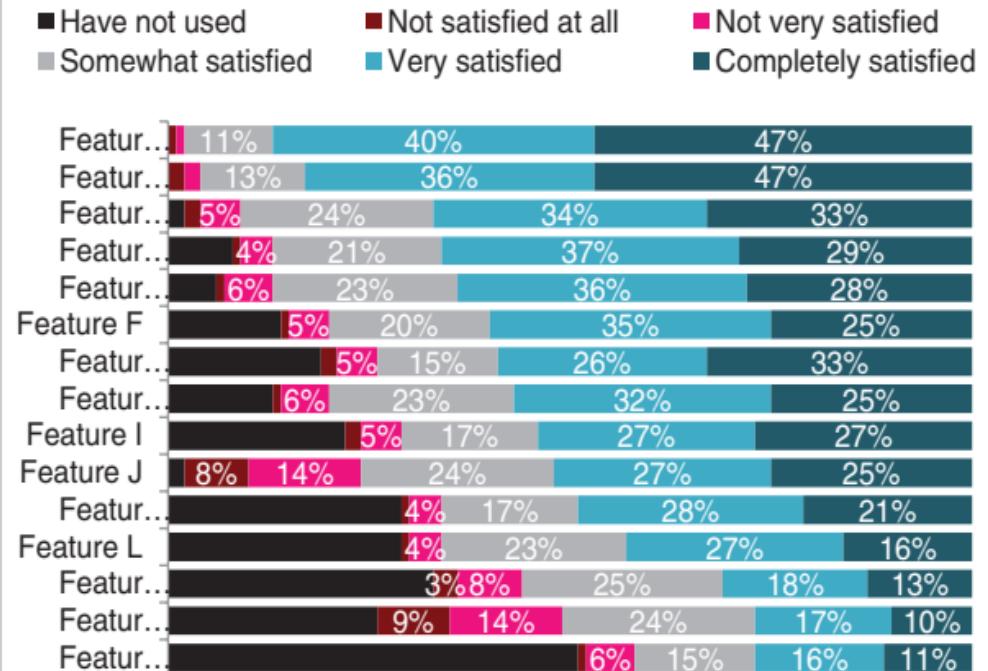


# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.

User Satisfaction

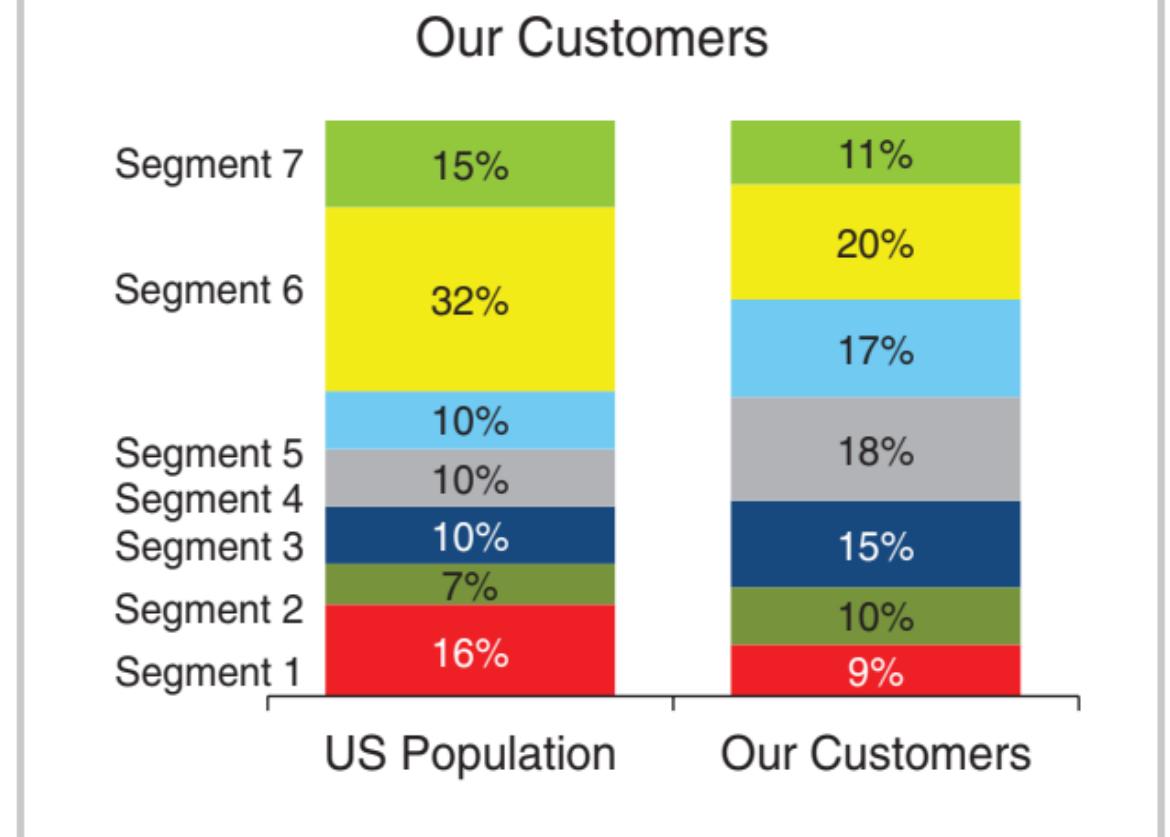




# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.

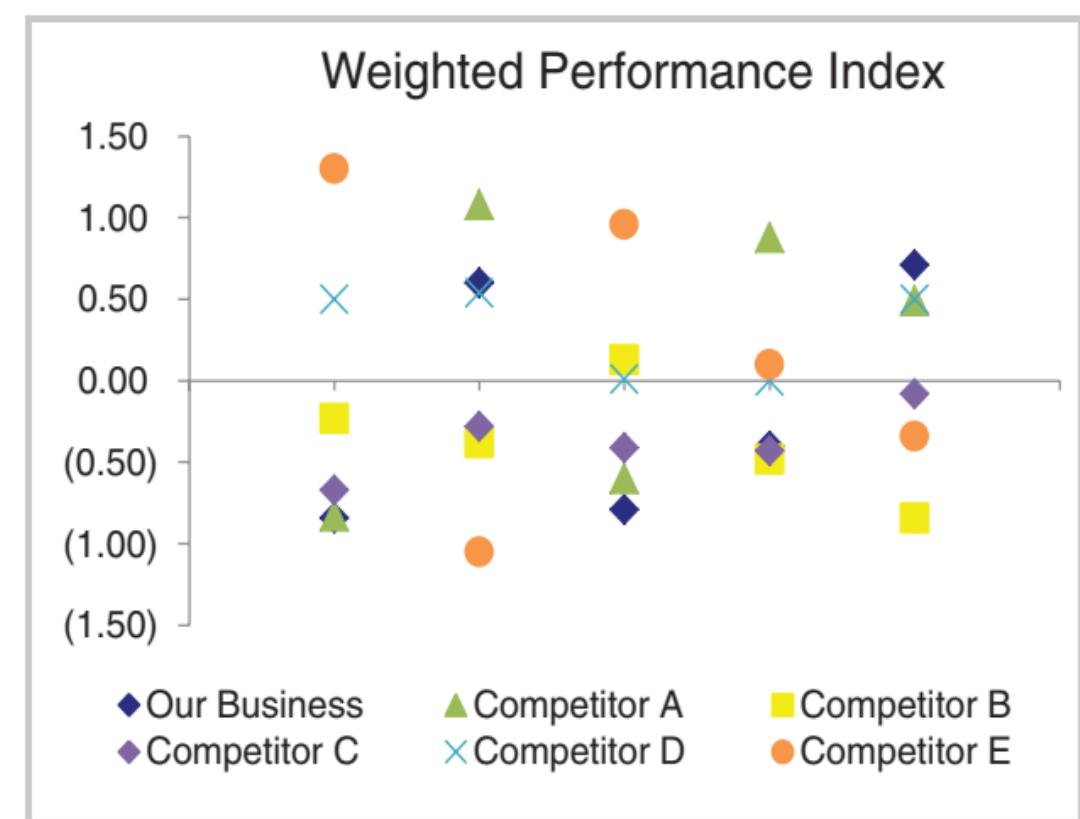




# Introduction

Bad graphs are everywhere:

- Nobody sets out to make a bad graph. But it happens. Again, and again.
- At every company throughout all industries and by all types of people.





# Introduction

- We aren't naturally good at storytelling with data



## Ticket Trend

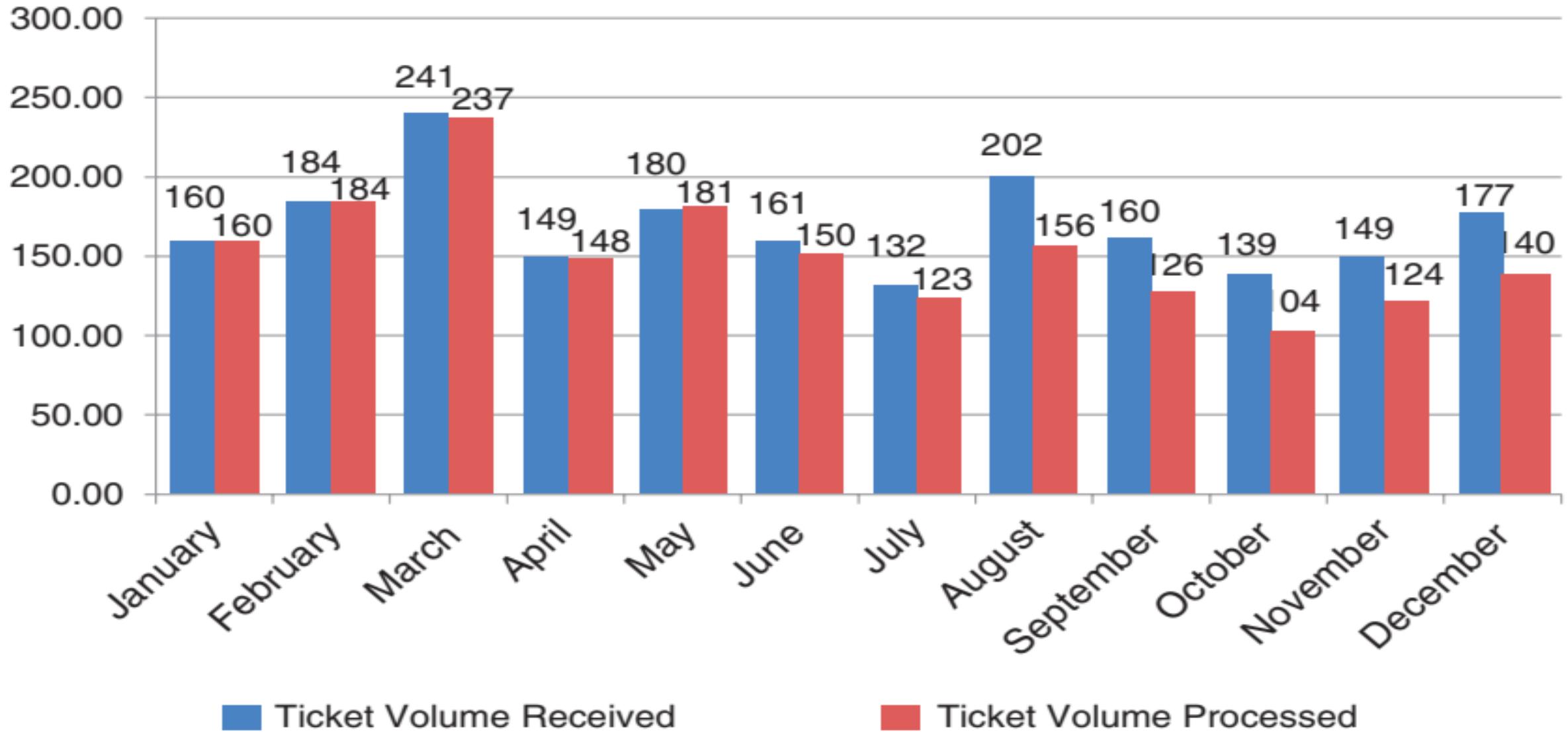


FIGURE 0.2 Example 1 (before): showing data

# Please approve the hire of 2 FTEs

to backfill those who quit in the past year

## Ticket volume over time



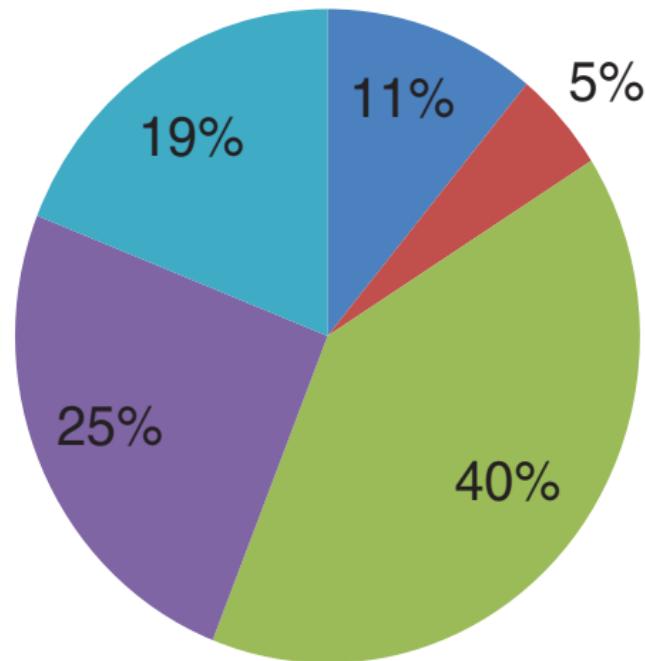
Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

**FIGURE 0.3** Example 1 (after): storytelling with data

# Survey Results

PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

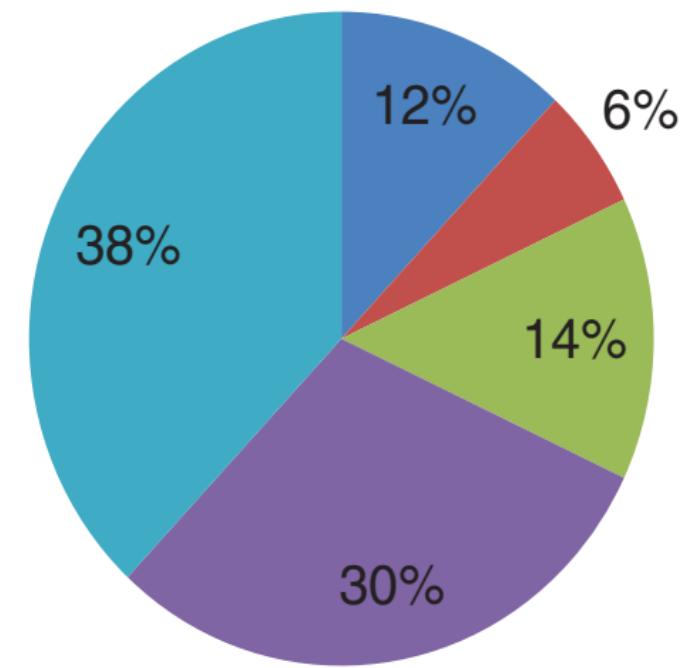
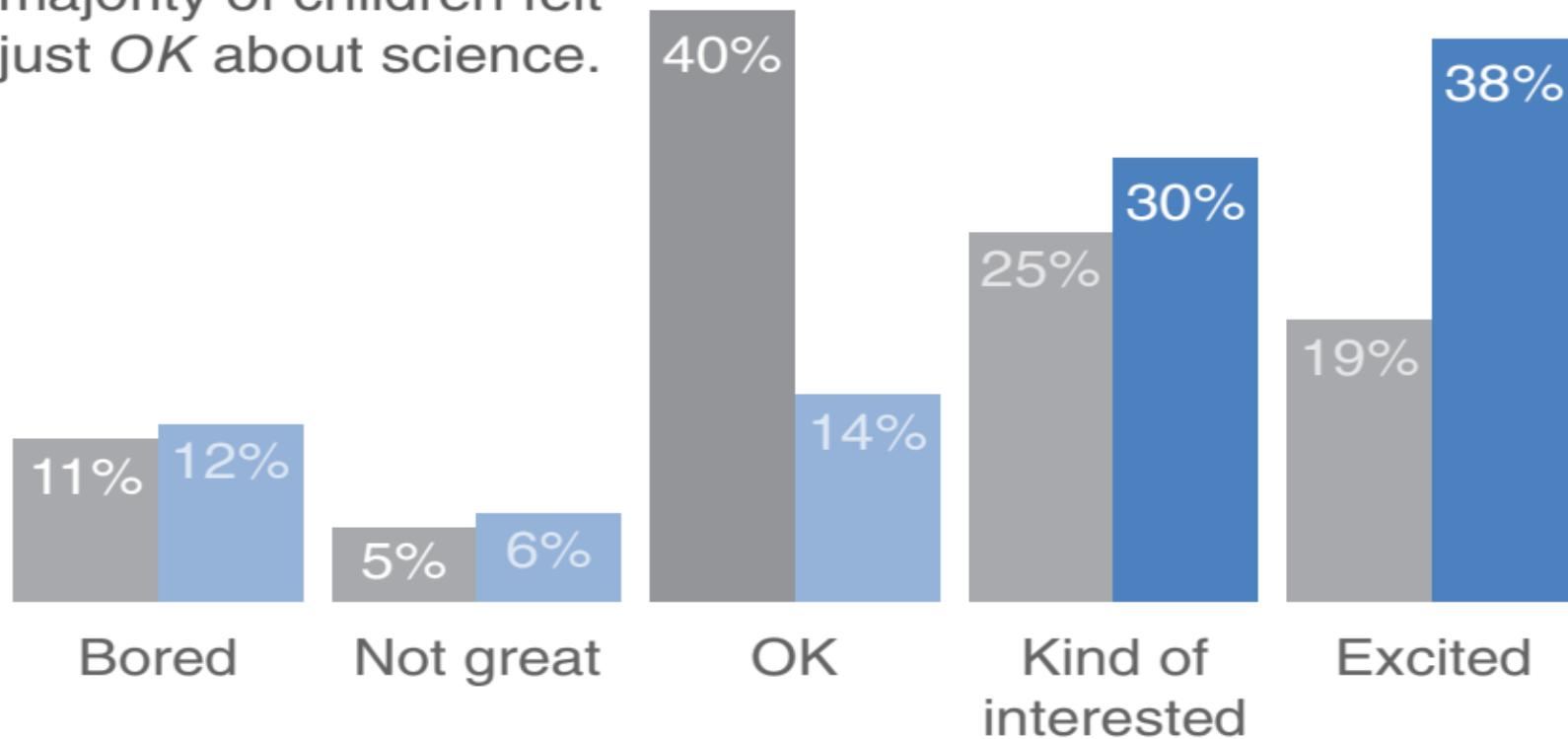


FIGURE 0.4 Example 2 (before): showing data

# Pilot program was a success

How do you feel about science?

**BEFORE** program, the majority of children felt just *OK* about science.



**AFTER** program,  
more children  
were *Kind of  
interested &  
Excited* about  
science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

**FIGURE 0.5** Example 2 (after): storytelling with data

## Average Retail Product Price per Year

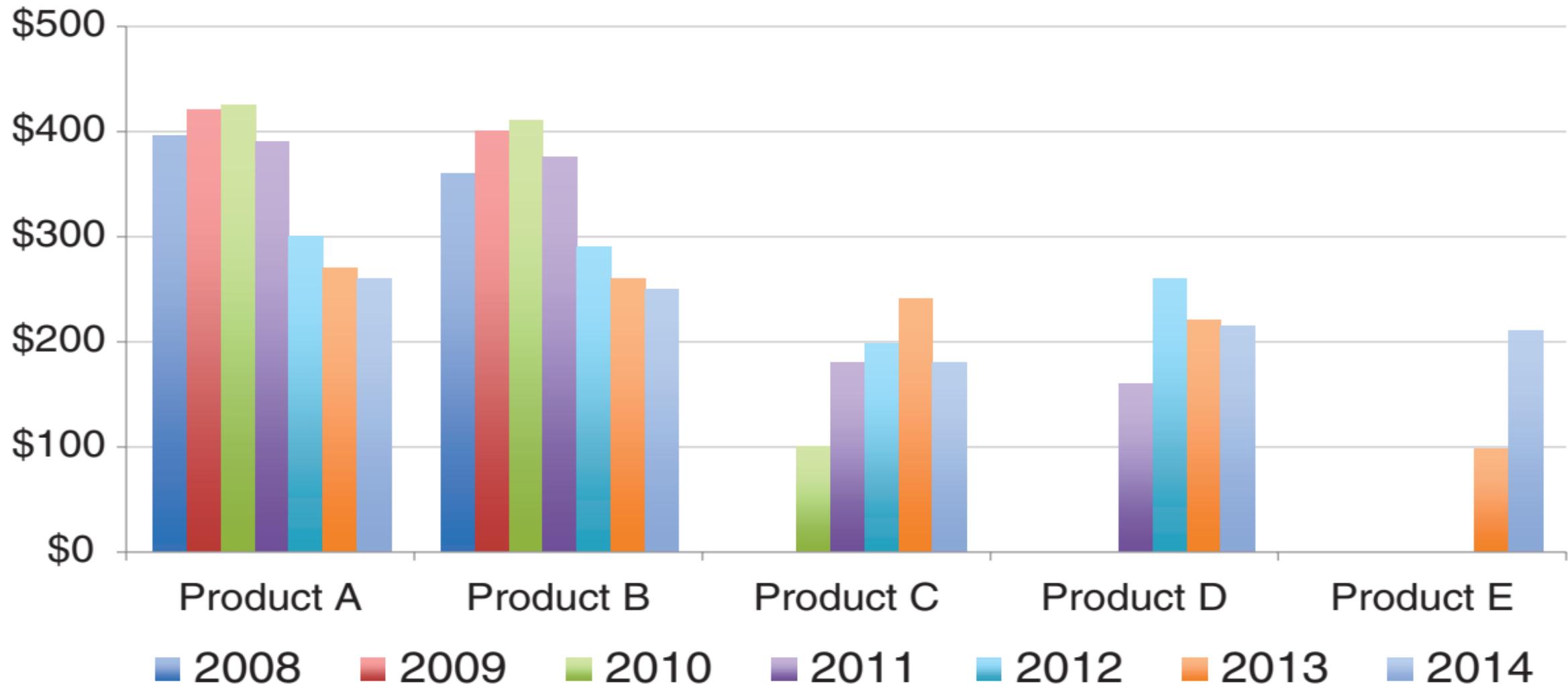


FIGURE 0.6 Example 3 (before): showing data

To be competitive, we recommend introducing our product *below* the \$223 average price point in the \$150–\$200 range

Retail price over time by product

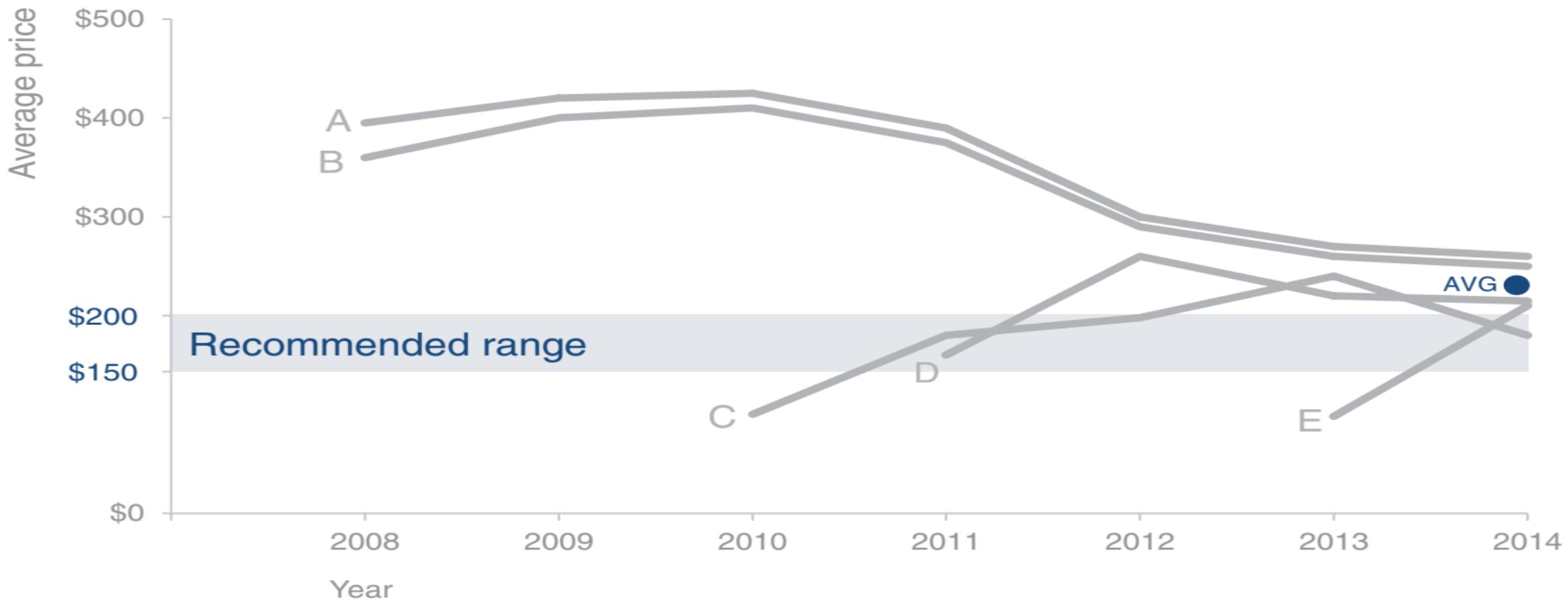


FIGURE 0.7 Example 3 (after): storytelling with data

# CORONAVIRUS

## STATS

As declared by WHO on March 11, 2020,  
COVID-19 has become a Pandemic



COUNTRY/REGION:



CONFIRMED



RECOVERED



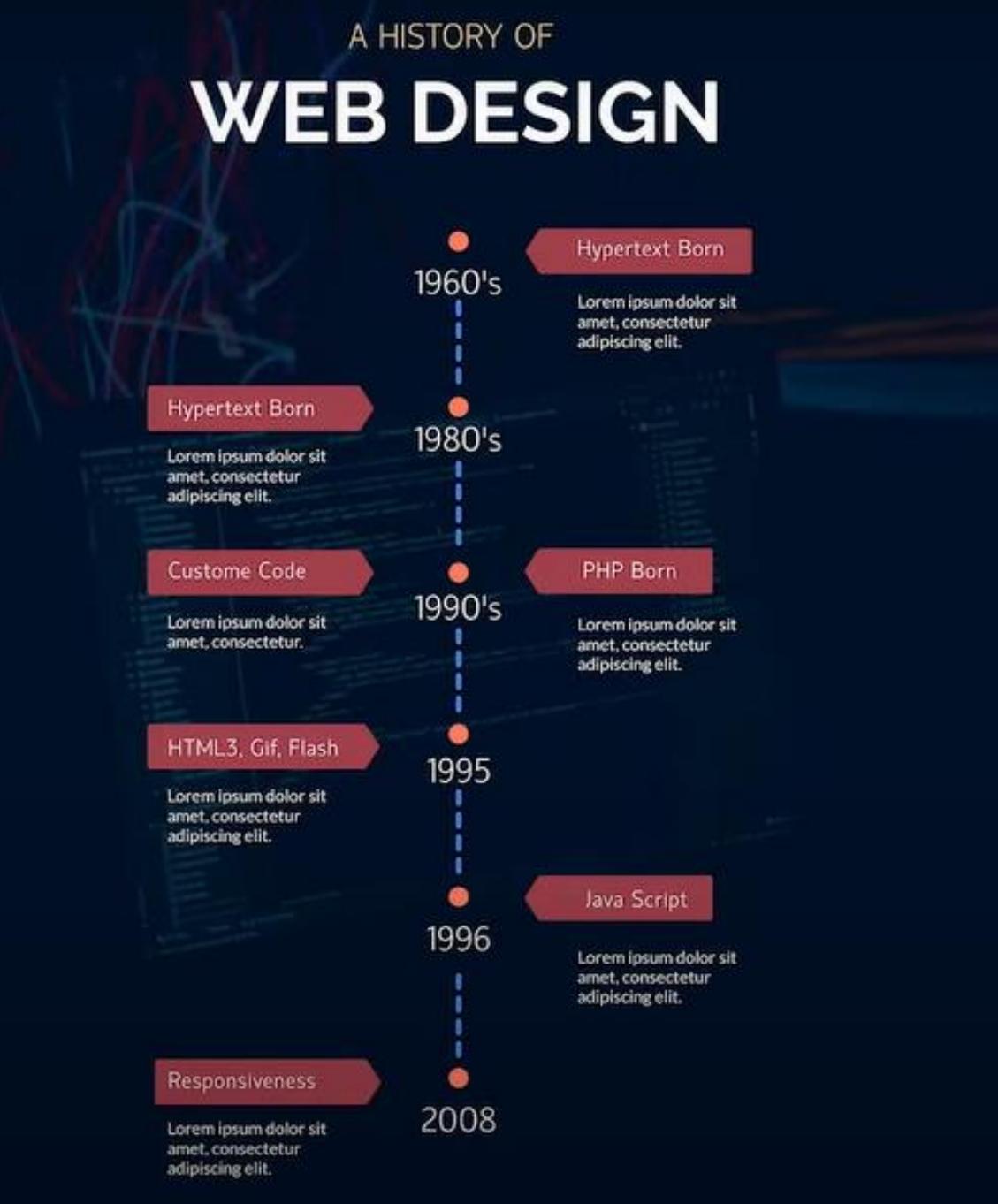
22,837

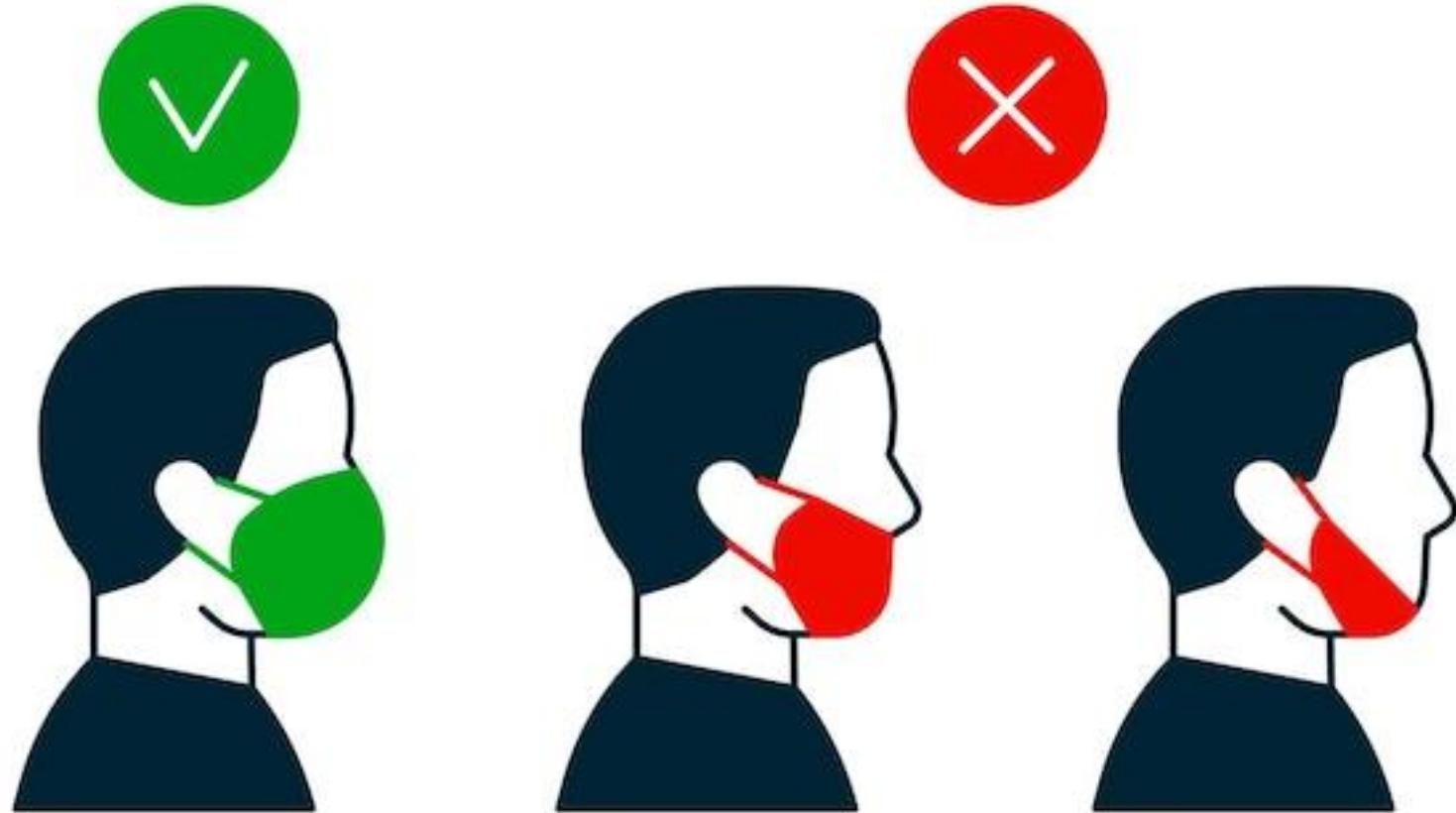
DEATHS



16,523

132,547







# Contents

1

Introduction

2

The Importance of Context

3

Choosing the Effective Visual



# DATA VISUALIZATION



# UNIT - I

# How you'll learn to tell stories with data



## STEP 1

### CONTEXT

Understand the Context



## STEP 2

### DISPLAY

Choose an appropriate visual display



## STEP 3

### CLUTTER

Eliminate Clutter



## STEP 4

### ATTENTION

Focus attention where you want it



## STEP 5

### THINKING

Think like a designer



## STEP 6

### TELL A STORY

Tell a story to audience





# Contents

1

Introduction

2

The Importance of Context

3

Choosing the Effective Visual



# The Importance of Context

- Success in data visualization does not start with data visualization.
- Rather, before you begin down the path of creating a data visualization or communication, attention and time should be paid to **understanding the context** for the need to communicate.





# The Importance of Context

## Exploratory vs. explanatory analysis:

- **Exploratory** analysis is what you do to understand the data and figure out what might be noteworthy or interesting to highlight to others.
- When we do exploratory analysis, it's like hunting for pearls in oysters.
- We might have to open 100 oysters to find perhaps two pearls.
- When we're at the point of communicating our analysis to our audience, we really want to be in the **explanatory** space, meaning you have a specific thing you want to explain, a specific story you want to tell—probably about those two pearls.



# The Importance of Context

## Exploratory analysis:

- Suppose researchers want to explore a new, mysterious disease that has recently emerged. They don't have much prior information about this disease. In this case, they might conduct exploratory research.





# The Importance of Context

- They could start by collecting data from affected patients, conducting interviews with them, and reviewing existing medical records.
- The goal here is to gain a better understanding of the disease, its symptoms, potential causes, and any patterns that might emerge.





# The Importance of Context

## Explanatory analysis:

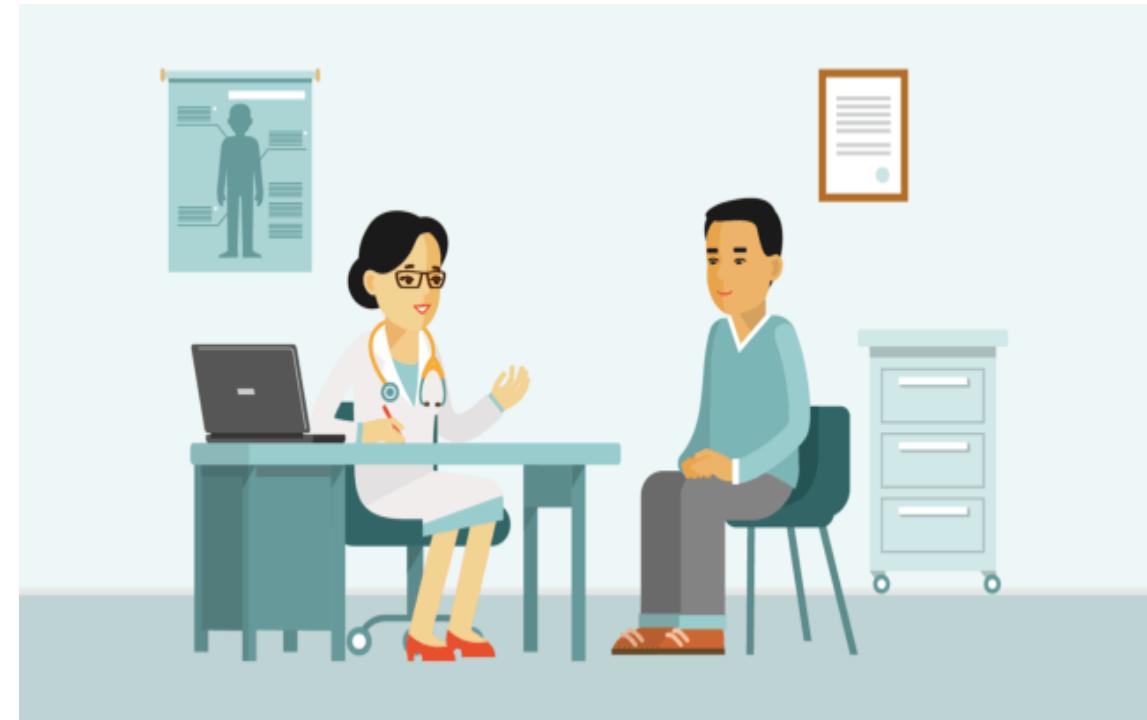
- The researchers have identified a potential cause for the disease, such as exposure to a certain environmental toxin. They now want to establish a causal relationship and explain how this toxin leads to the disease.





# The Importance of Context

- To do this, they design a controlled experiment, exposing laboratory animals to the toxin and monitoring their health over time. This explanatory research aims to provide clear evidence of causation by controlling variables and using statistical analysis to establish a cause-and-effect relationship.





# The Importance of Context

**Who, what, and how:**

- When it comes to explanatory analysis, there are a few things to think about and be extremely clear on before visualizing any data or creating content.
- **To whom are you communicating?** It is important to have a good understanding of who your audience is and how they perceive you.





# The Importance of Context

- What do you want your audience to know or do?
- You should be clear how you want your audience to act and consider how you will communicate to them and the overall tone that you want to set for your communication.
- It's only after you can concisely answer these first two questions that you're ready to move forward with How can you use data to help make your point?

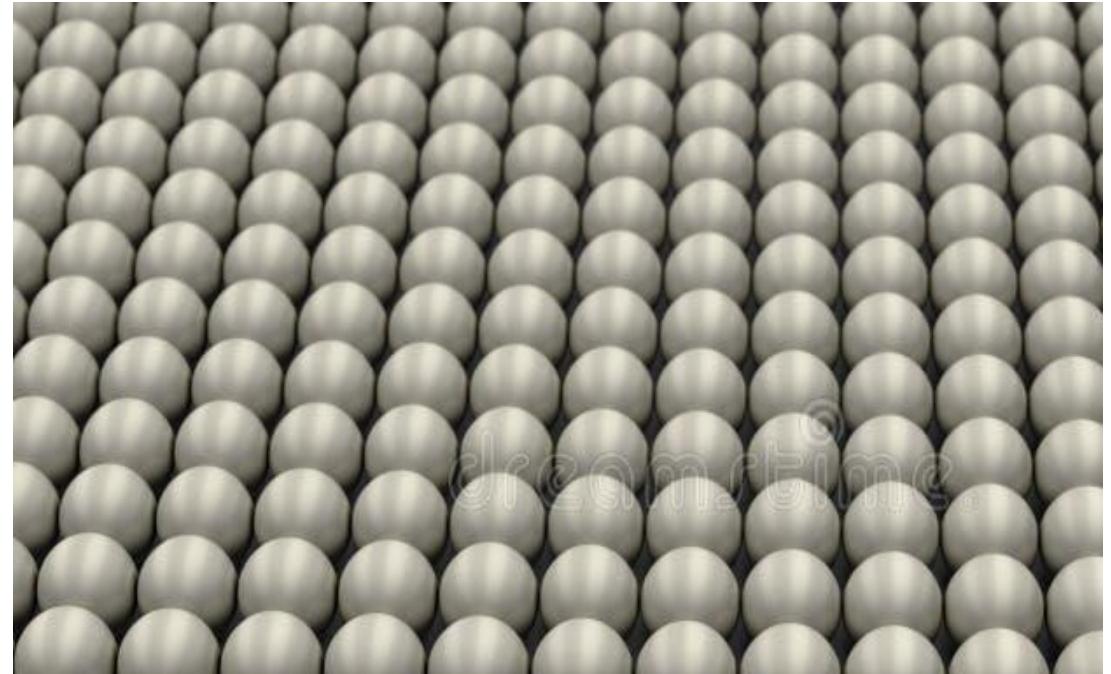


# The Importance of Context

Who:

Your audience:

- The **more specific** you can be about who your audience is, the better position you will be in for successful communication.





# The Importance of Context

Who:

Your audience:

- Avoid general audiences, such as “internal and external stakeholders” by trying to communicate to too many different people with disparate needs at once, you put yourself in a position where you can’t communicate to any one of them.

## Stakeholders





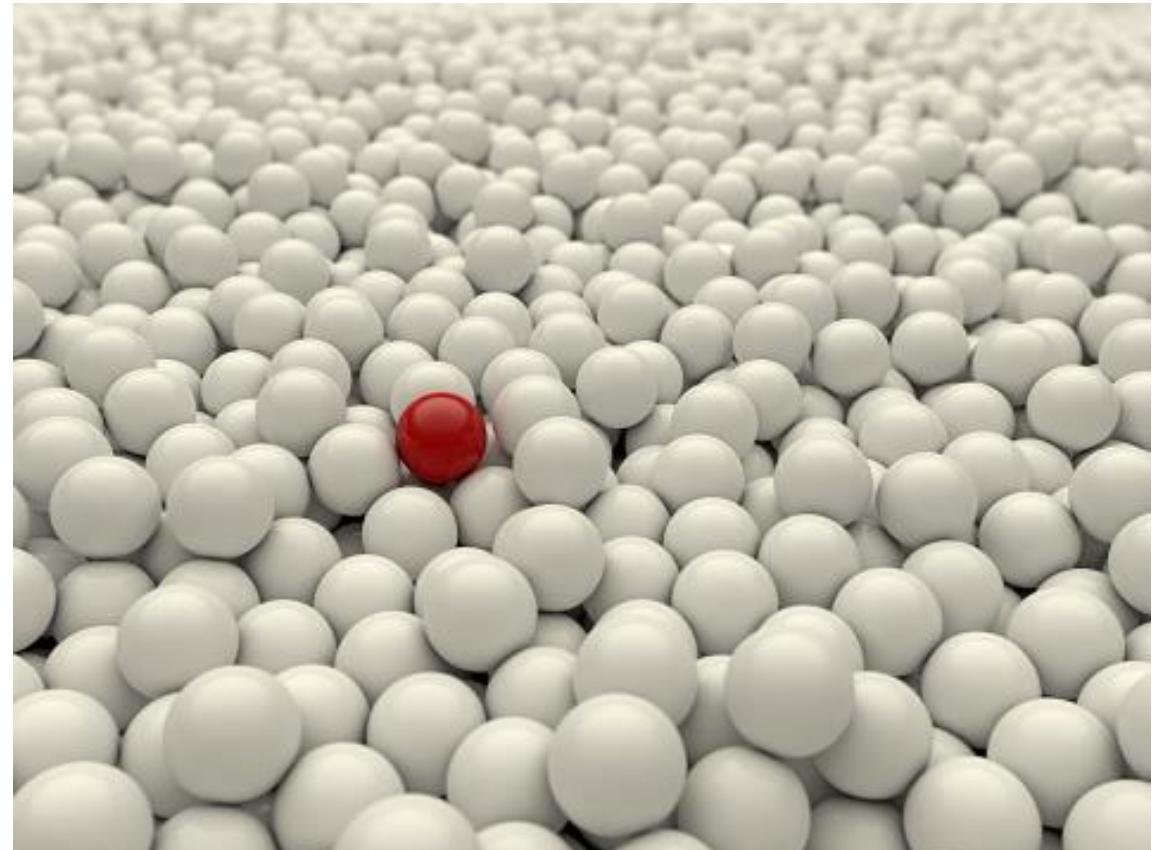
# The Importance of Context

Who:

Your audience:

- Sometimes this means creating different communications for different audiences.

Identifying the **decision maker** is one way of narrowing your audience





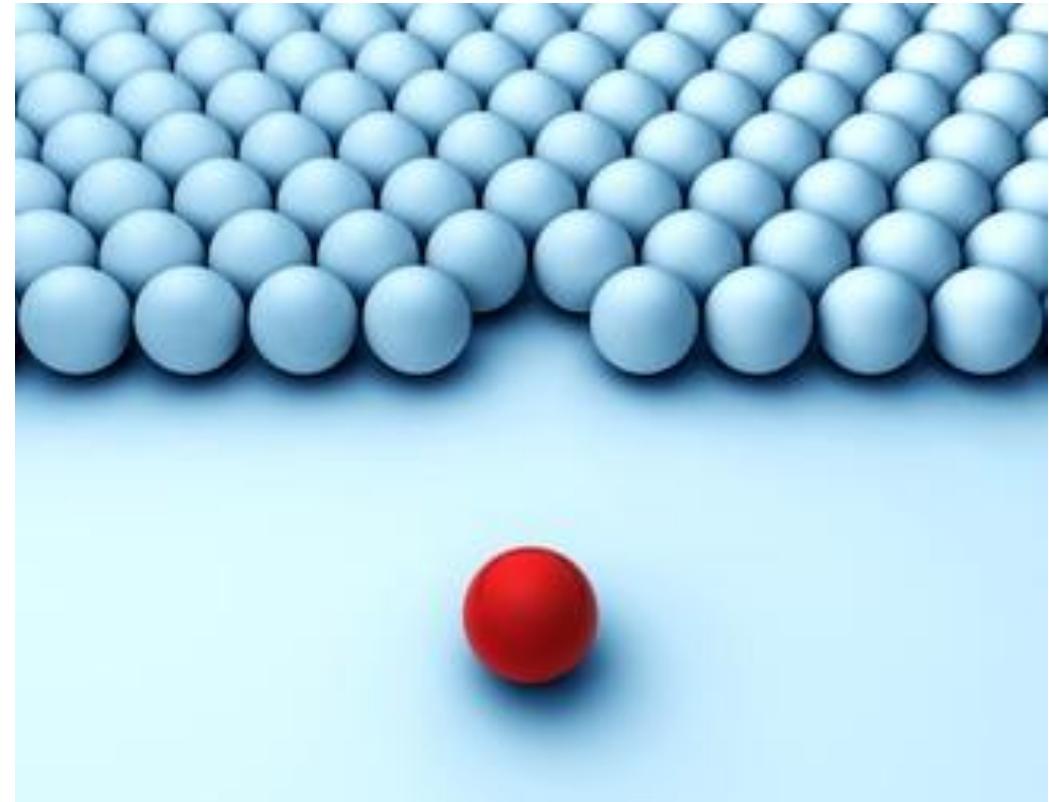
# The Importance of Context

Who:

Your audience:

- Sometimes this means creating different communications for different audiences.

Identifying the **decision maker** is one way of narrowing your audience



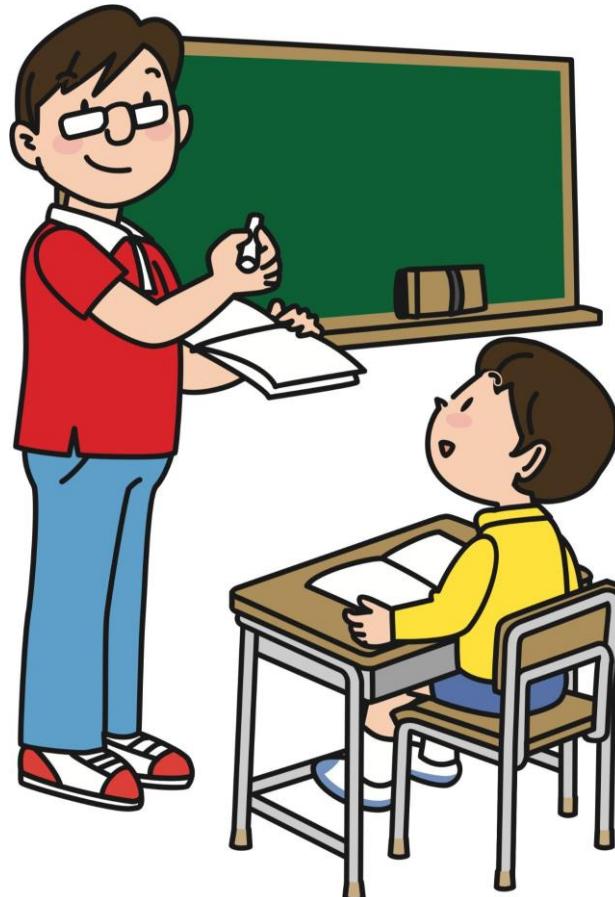


# The Importance of Context

Who:

You:

- It's also helpful to think about the relationship that you have with your audience and how you expect that they will perceive you.





# The Importance of Context

Who:

You:

- Will you be encountering each other for the first time through this communication, or do you have an established relationship? Do they already trust you as an expert, or do you need to work to establish credibility?





# The Importance of Context

What:

Action:

- What do you need your audience to know or do? This is the point where you think through how to make what you communicate relevant for your audience and form a clear understanding of why they should care about what you say.



# The Importance of Context

What:

Mechanism:

- How will you communicate to your audience?
- We can think of the communication mechanism along a continuum, with live presentation at the left and a written document or email at the right, as shown in Figure.

LIVE PRESENTATION ..... WRITTEN DOC OR EMAIL

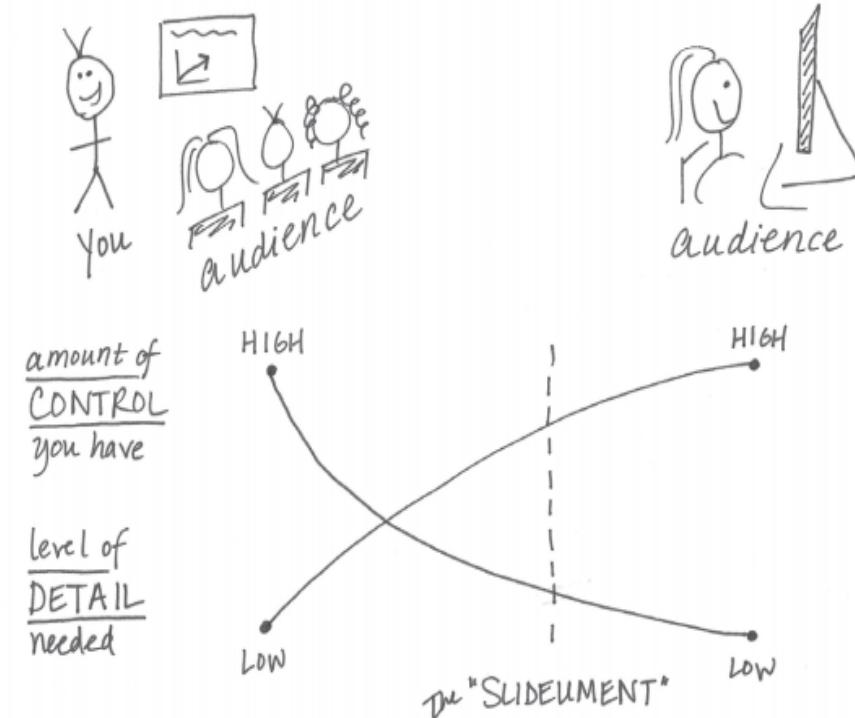


FIGURE 1.1 Communication mechanism continuum



# The Importance of Context

What:

Mechanism:

- At the left, with a live presentation, you (the presenter) are in full control. You determine what the audience sees and when they see it. You can respond to visual cues to speed up, slow down, or go into a particular point in detail.

LIVE PRESENTATION ..... WRITTEN DOC OR EMAIL

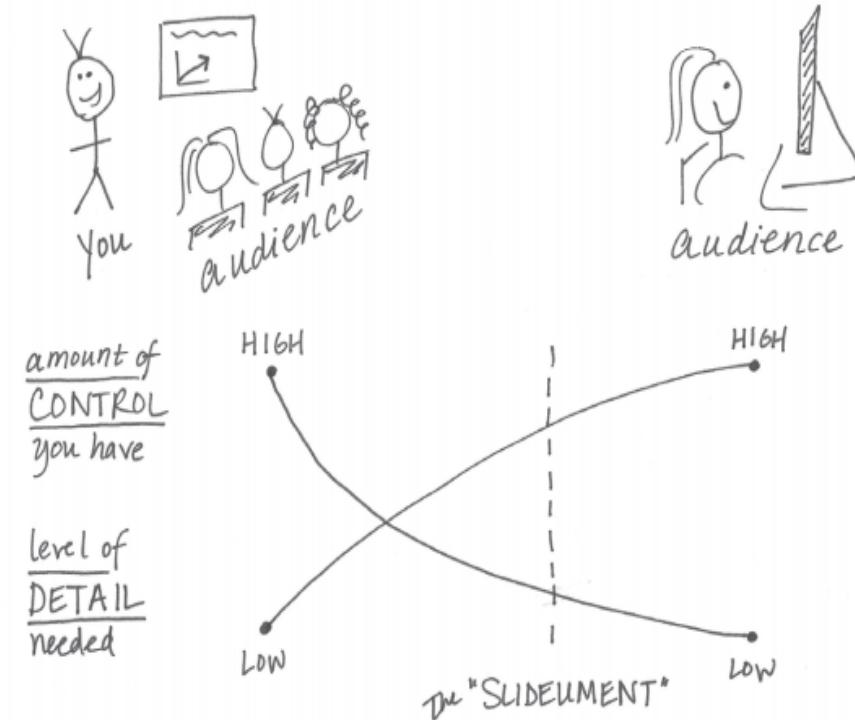


FIGURE 1.1 Communication mechanism continuum



# The Importance of Context

What:

Mechanism:

- At the right side of the spectrum, with a written document or email, you have less control. In this case, the audience is in control of how they consume the information.

LIVE PRESENTATION ..... WRITTEN DOC OR EMAIL

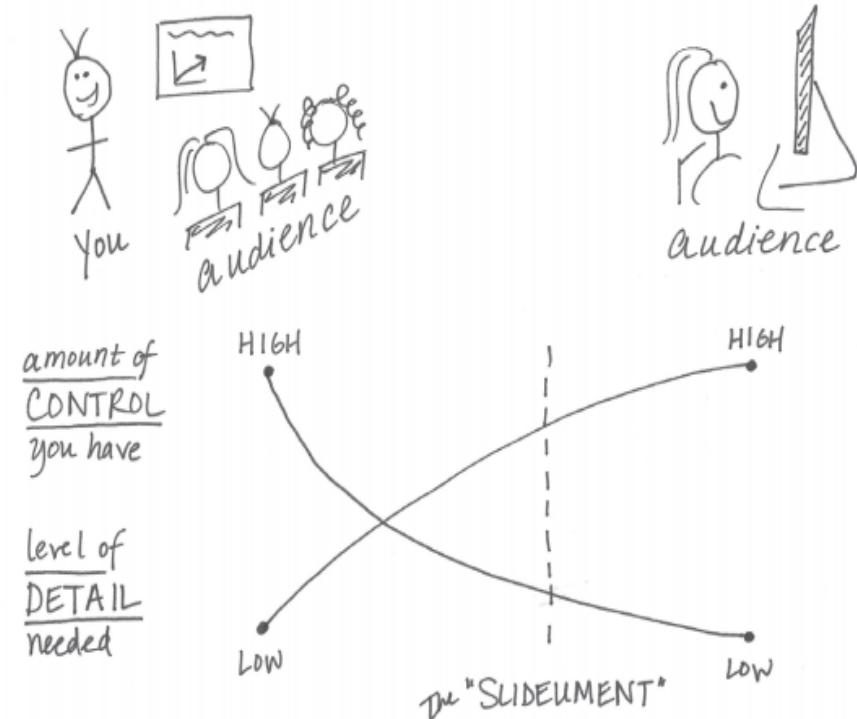


FIGURE 1.1 Communication mechanism continuum



# The Importance of Context

What:

Mechanism:

- This gives rise to the slideument, a single document that's meant to solve both needs.

**SLIDES ARE  
NOT  
DOCUMENTS**



# The Importance of Context

What:

Tone:

- What tone do you want your communication to set?
- Are you celebrating a success?
- Trying to light a fire to drive action?
- Is the topic lighthearted or serious?

How:

- What data is available that will help make my point?
- Data becomes supporting evidence of the story you will build and tell.



# Contents

1

Introduction

2

The Importance of Context

3

Choosing the Effective Visual



# DATA VISUALIZATION



# UNIT - I

# How you'll learn to tell stories with data



## STEP 1

### CONTEXT

Understand the Context



## STEP 2

### DISPLAY

Choose an appropriate visual display



## STEP 3

### CLUTTER

Eliminate Clutter



## STEP 4

### ATTENTION

Focus attention where you want it



## STEP 5

### THINKING

Think like a designer



## STEP 6

### TELL A STORY

Tell a story to audience





# Contents

1

Introduction

2

The Importance of Context

3

Choosing the Effective Visual



# Choosing the Effective Visual

- There are many different graphs and other types of visual displays of information.
- Choosing an effective visual, whether it's for a presentation, marketing campaign, website, or any other purpose, is crucial for conveying your message and engaging your audience.

**So Many  
Graphs & Displays**



# Choosing the Effective Visual



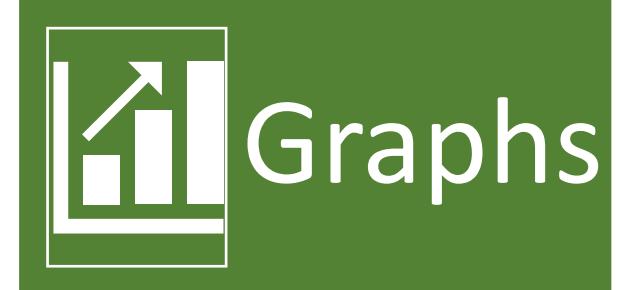
# Choosing the Effective Visual

- Column
- Line
- Pie
- Bar
- Area
- X Y (Scatter)
- Map
- Stock
- Surface
- Radar
- Treemap
- Sunburst
- Histogram
- Box & Whisker
- Waterfall
- Funnel
- Combo

- List
- Process
- Cycle
- Hierarchy
- Relationship
- Matrix
- Pyramid
- Picture



## Visual Types





- Simple Text
- Tables
- Graphs



- When you have just a number or two to share, simple text can be a great way to communicate.

**91%**

Simple text

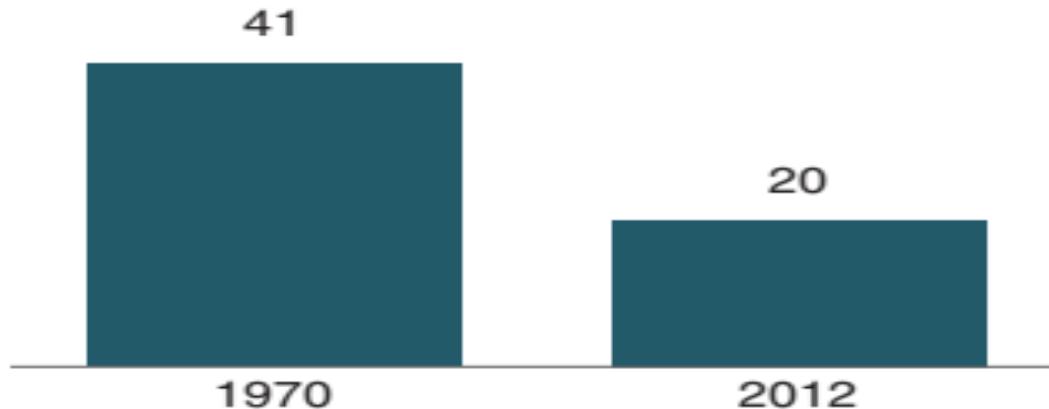


- Simple Text
- Tables
- Graphs



## Children with a "Traditional" Stay-at-Home Mother

*% of children with a married stay-at-home mother with a working husband*



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from **PEW RESEARCH CENTER**

**FIGURE 2.2** Stay-at-home moms original graph



- Simple Text
- Tables
- Graphs



20%

of children had a  
**traditional stay-at-home mom**  
in 2012, compared to 41% in 1970

FIGURE 2.3 Stay-at-home moms simple text makeover



- Simple Text
- Tables →
- Graphs

- Tables interact with our verbal system, which means that we read them.
- When I have a table in front of me, I typically have my index finger out: I'm reading across rows and down columns or I'm comparing values.

### Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ



- Simple Text
- Tables →
- Graphs

- Tables interact with our verbal system, which means that we read them.
- When I have a table in front of me, I typically have my index finger out: I'm reading across rows and down columns or I'm comparing values.

### Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ



- Simple Text
- Tables
- Graphs



- Tables interact with our verbal system, which means that we read them.
- When I have a table in front of me, I typically have my index finger out: I'm reading across rows and down columns or I'm comparing values.

### Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ



- Simple Text
- Tables
- Graphs



- Borders should be used to improve the legibility of your table.
- Think about pushing them to the background by making them grey or getting rid of them altogether.
- The data should be what stands out, not the borders.



- Simple Text
- Tables →
- Graphs

- let's shift our focus to a special case of tables:

Heatmap.

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%



- Simple Text
- Tables →
- Graphs

- In the left table, I find myself scanning across rows and down columns to get a sense of what I'm looking at, where numbers are higher or lower.

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%



- Simple Text
- Tables →
- Graphs

- We can use color saturation to provide visual cues, helping our eyes and brains more quickly target the potential points of interest. In the second iteration of the table on the right entitled “Heatmap,” the higher saturation of blue, the higher the number.

Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%



- Simple Text
- Tables
- Graphs

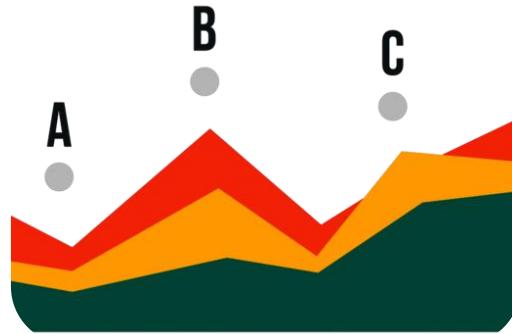


- Graphs interact with our visual system, which is faster at processing information. This means that a well-designed graph will typically get the information across more quickly than a well-designed table.





- Simple Text
- Tables
- **Graphs**



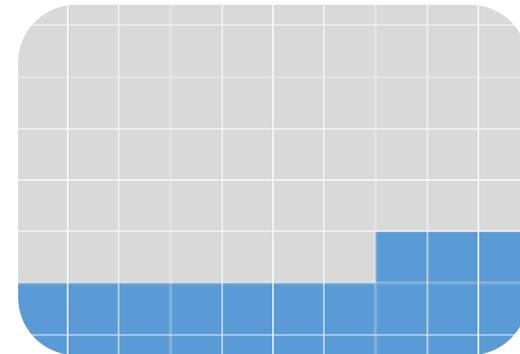
Points



Lines



Bars



Area



- Points
- Lines
- Bars
- Area



### Scatterplot:

- Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal x-axis and vertical y-axis to see whether and what relationship exists.
- For example, let's say that we manage a bus fleet and want to understand the relationship between miles driven and cost per mile.



- Points
- Lines
- Bars
- Area

Cost per mile by miles driven

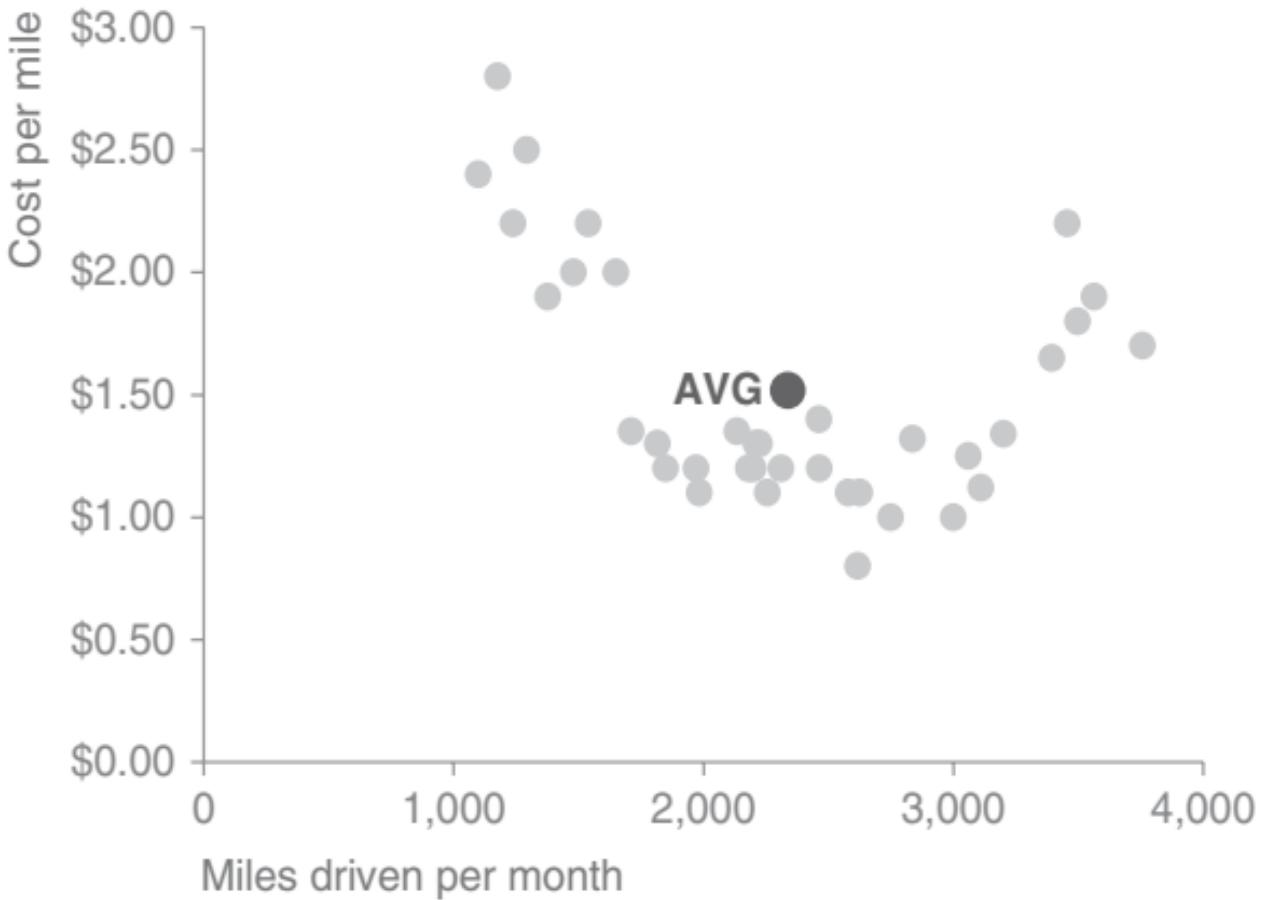


FIGURE 2.6 Scatterplot



- Points
- Lines
- Bars
- Area



Cost per mile by miles driven

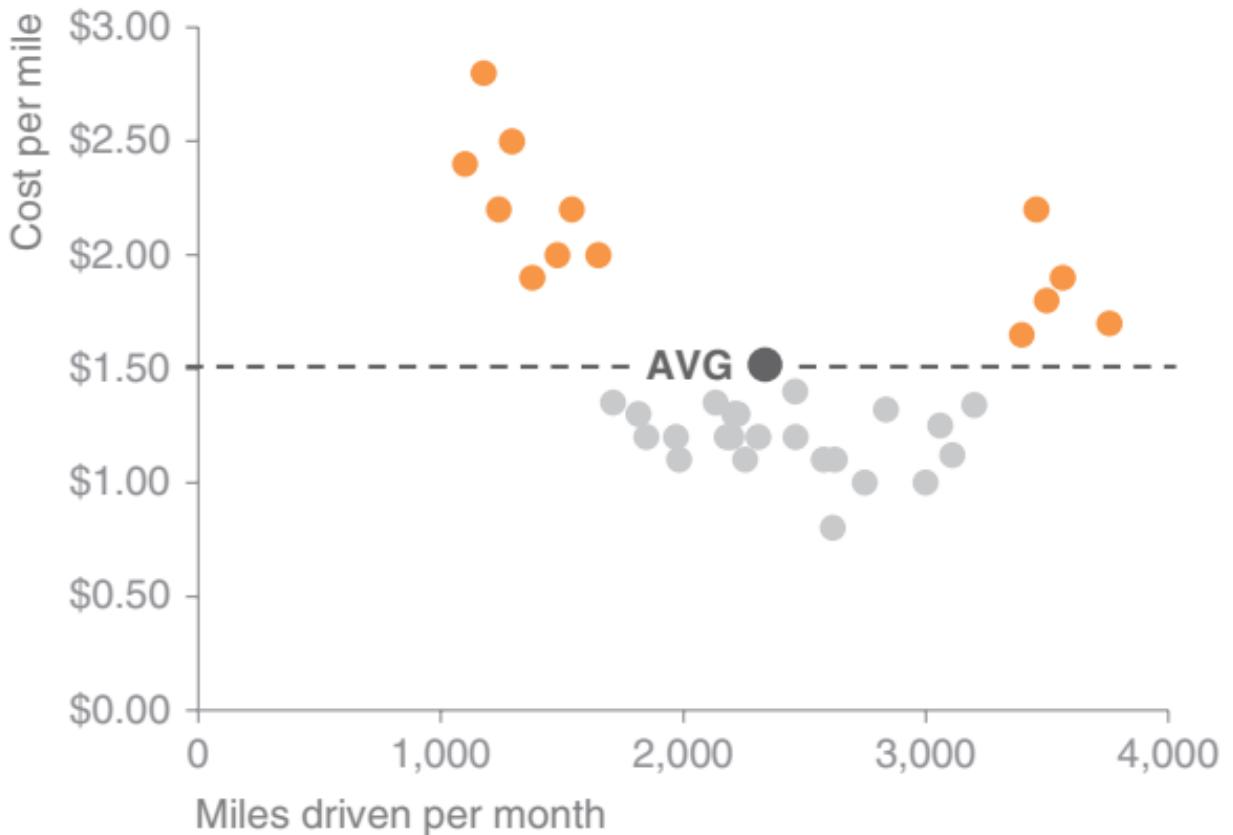


FIGURE 2.7 Modified scatterplot



- Points
- Lines
- Bars
- Area



Cost per mile by miles driven

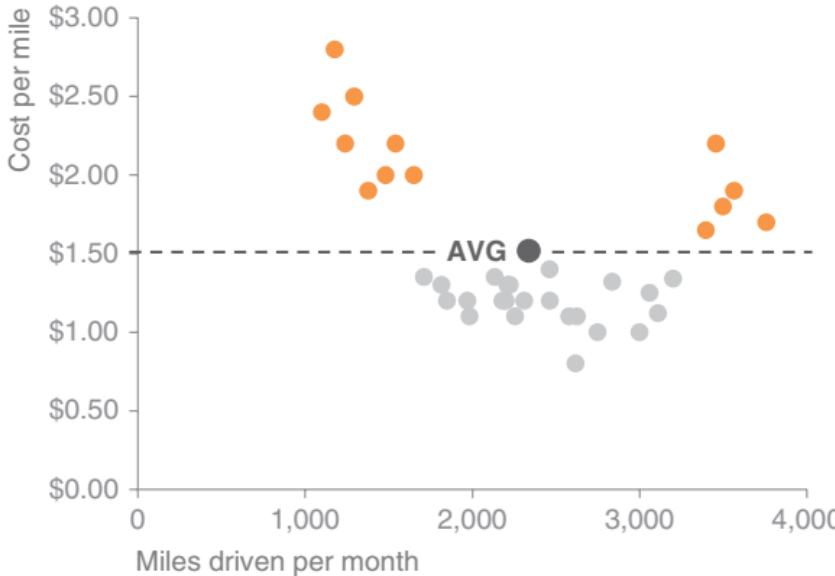


FIGURE 2.7 Modified scatterplot

- Observations such as cost per mile is higher than average when less than about 1,700 miles or more than about 3,300 miles were driven for the sample observed.



- Points
- Lines
- Bars
- Area



- Line graphs are most commonly used to plot continuous data.
- Because the points are physically connected via the line, it implies a connection between the points.
- There are two types:
  1. The standard line graph
  2. The slope graph.

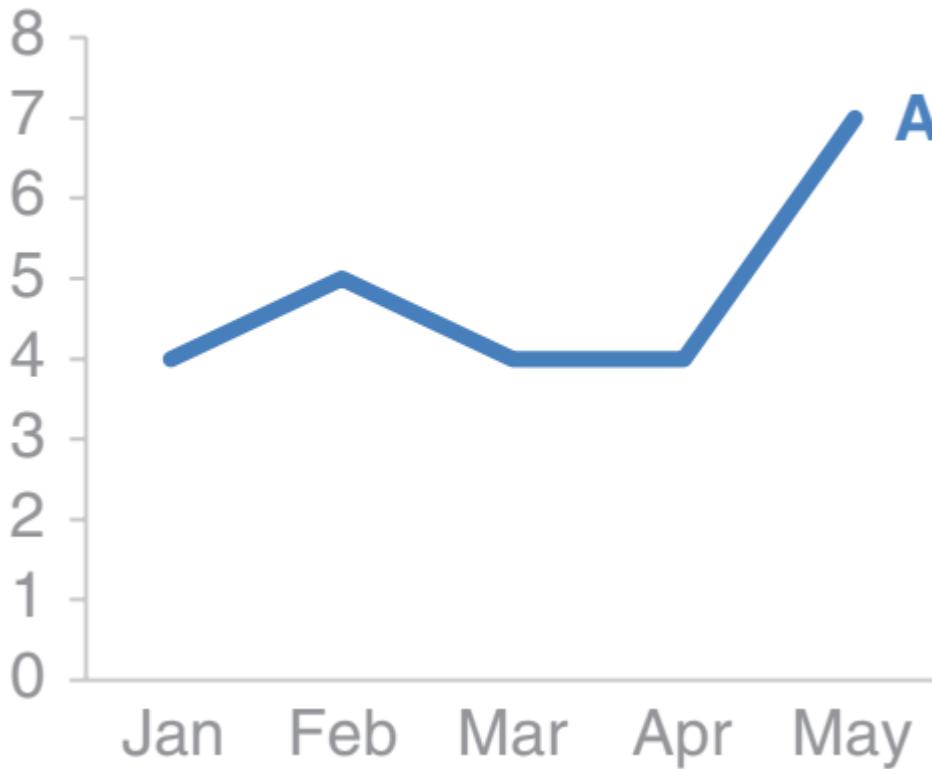


- Points
- Lines
- Bars
- Area



Line graph:

Single series



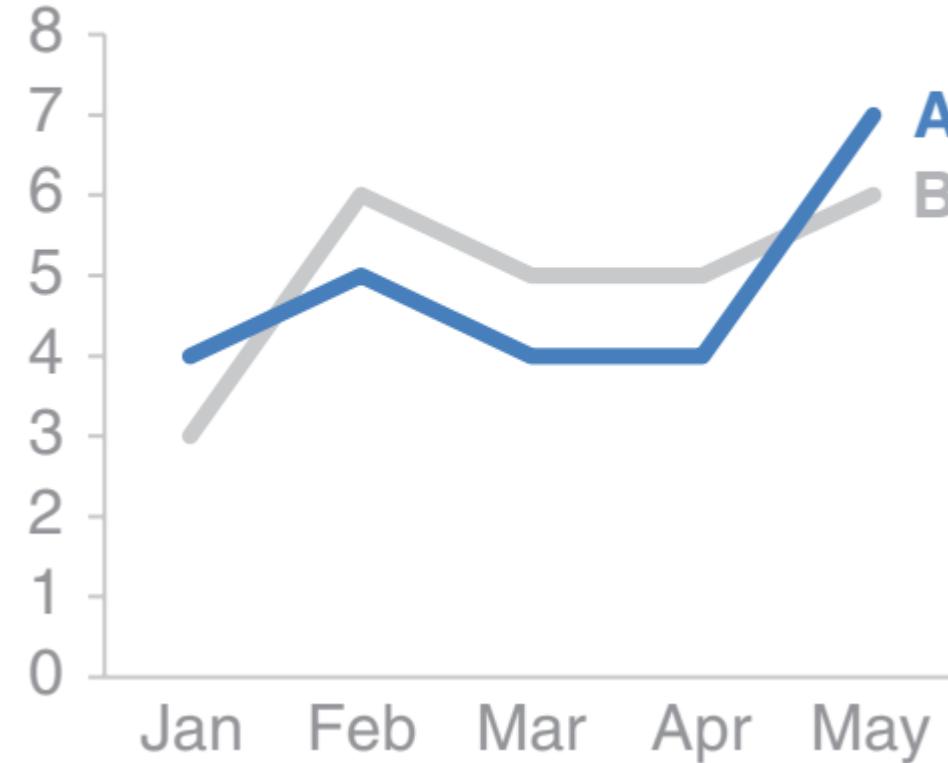


- Points
- Lines
- Bars
- Area



Line graph:

Two series



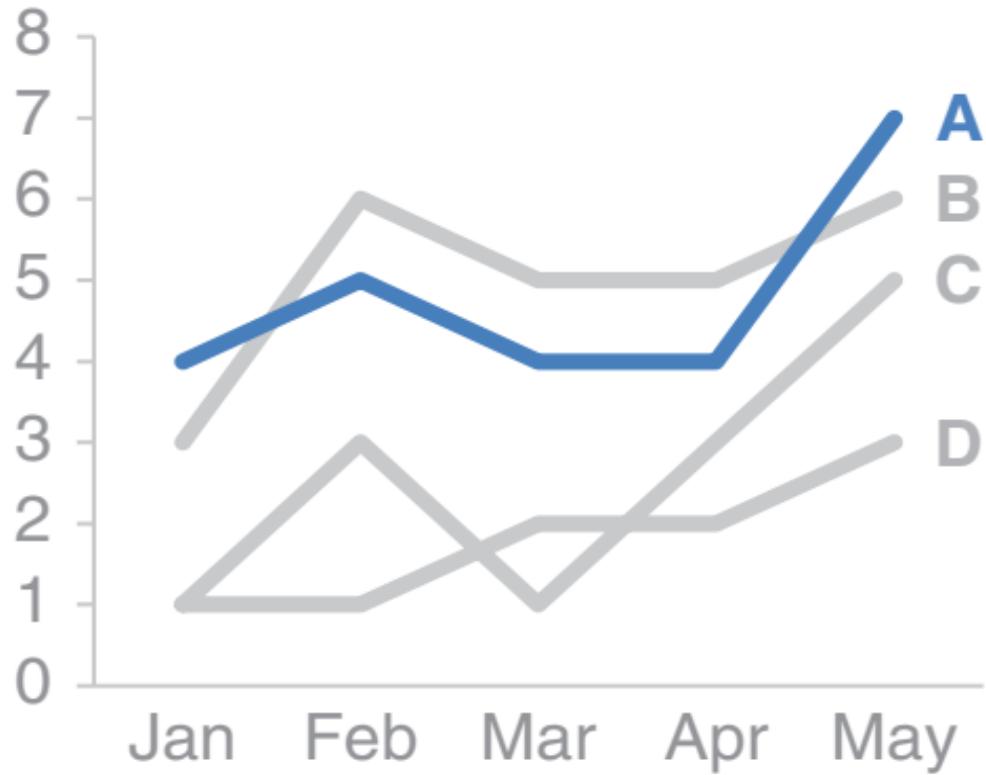


- Points
- Lines
- Bars
- Area



Line graph:

## Multiple series





- Points
- Lines
- Bars
- Area



Passport control wait time  
Past 13 months

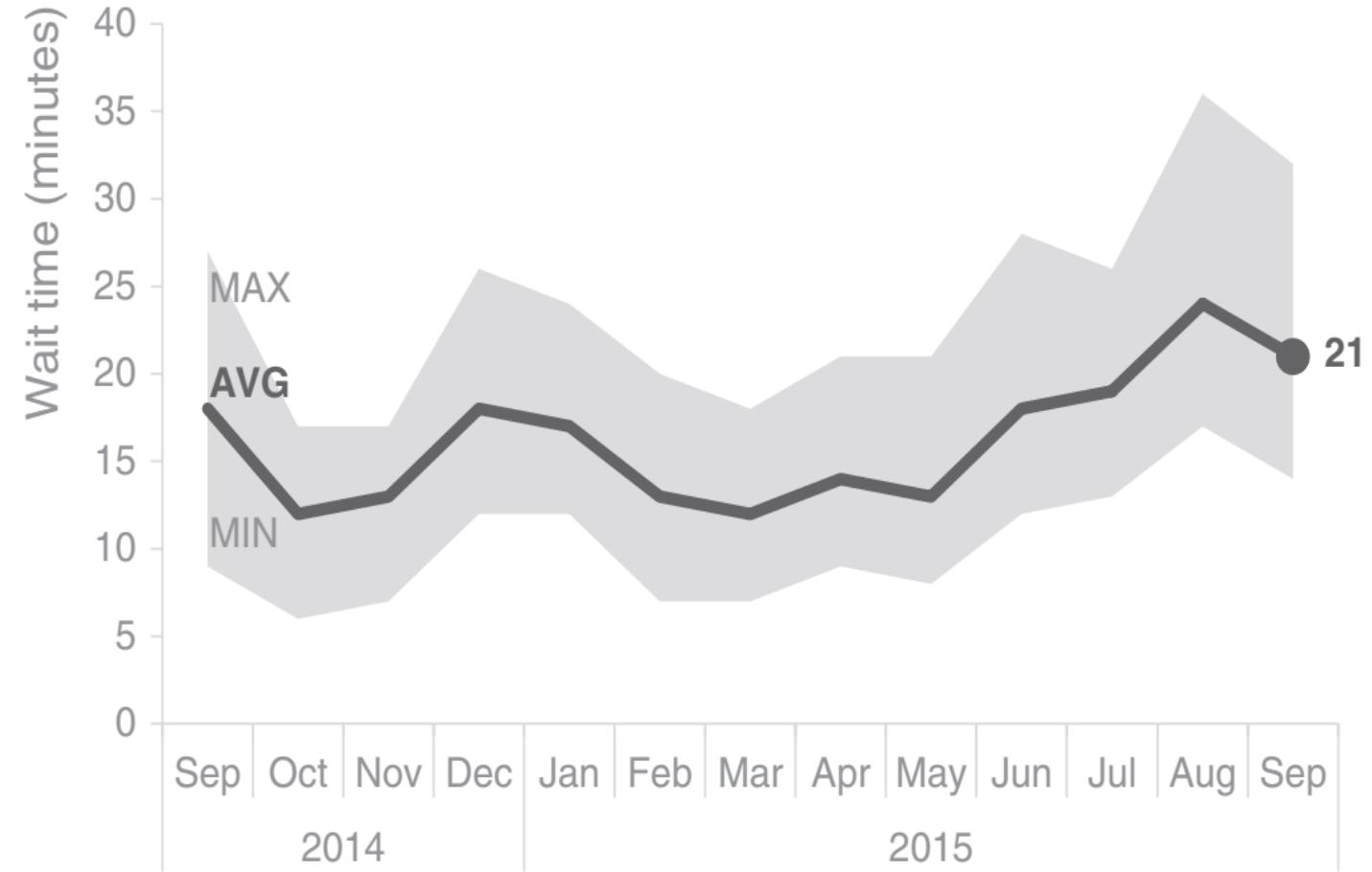


FIGURE 2.9 Showing average within a range in a line graph



- Points
- Lines
- Bars
- Area



### Slope graph:

- Slope graphs can be useful when you have two time periods.
- Points of comparison and want to quickly show relative increases and decreases.
- Differences across various categories between the two data points.



- Points
- Lines
- Bars
- Area



## Employee feedback over time

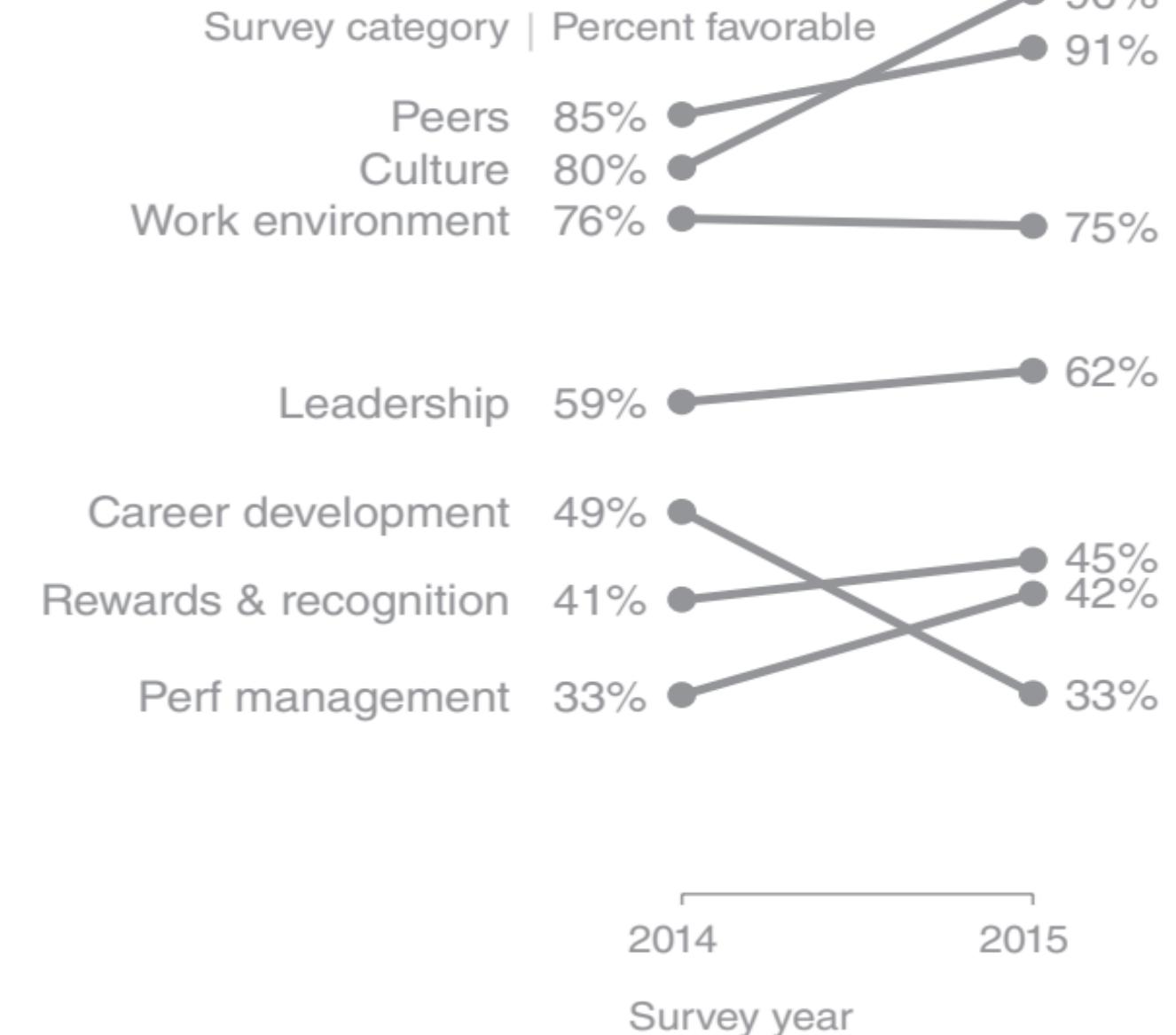


FIGURE 2.10 Slopegraph



- Points
- Lines
- Bars
- Area



## Employee feedback over time

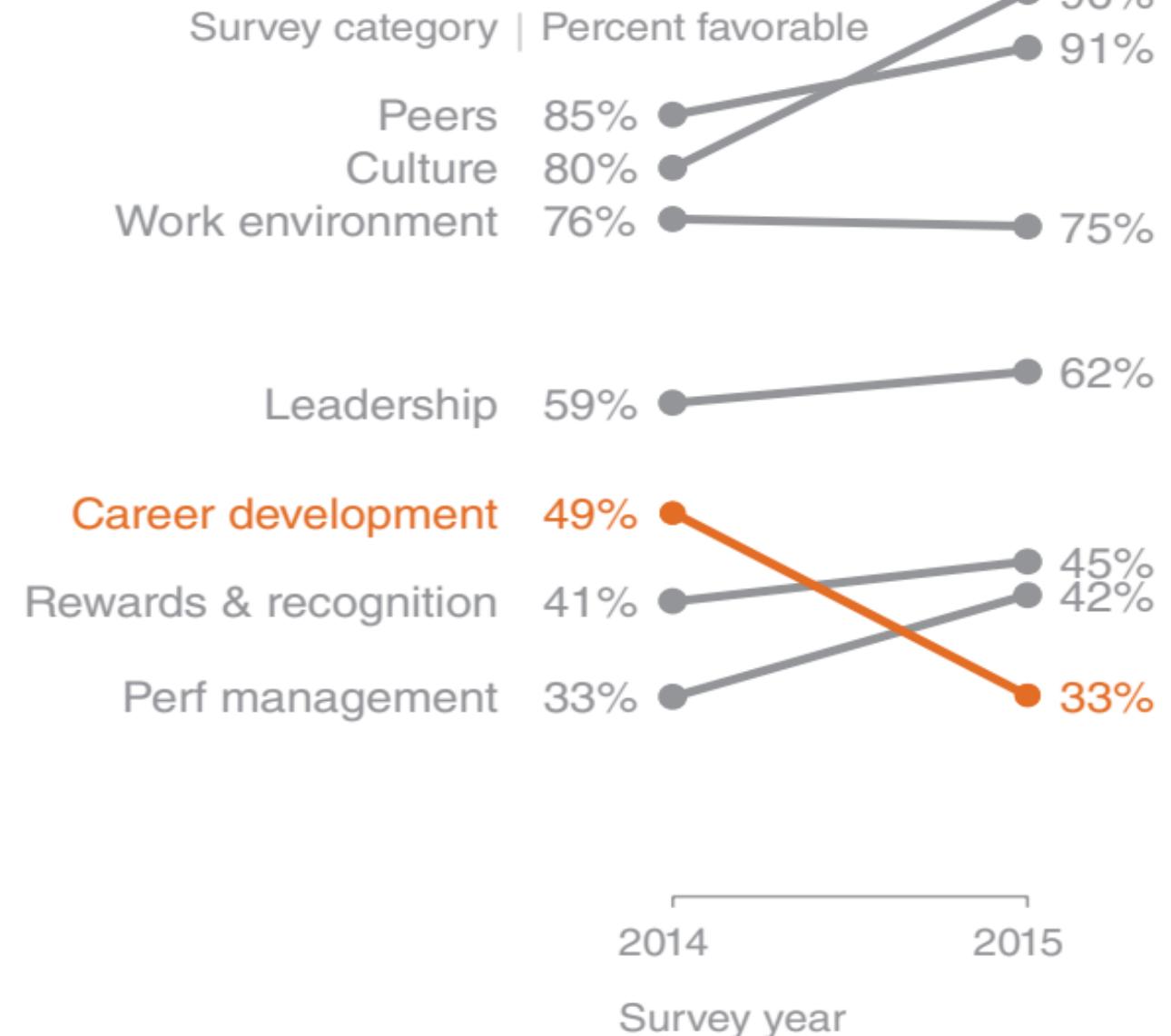


FIGURE 2.11 Modified slopegraph



- Points
- Lines
- Bars
- Area



- Bar charts are easy for our eyes to read. Our eyes compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and the incremental difference between categories.
- It is important that bar charts always have a zero baseline (where the x-axis crosses the y-axis at zero), otherwise you get a false visual comparison.



- Points
- Lines
- Bars
- Area

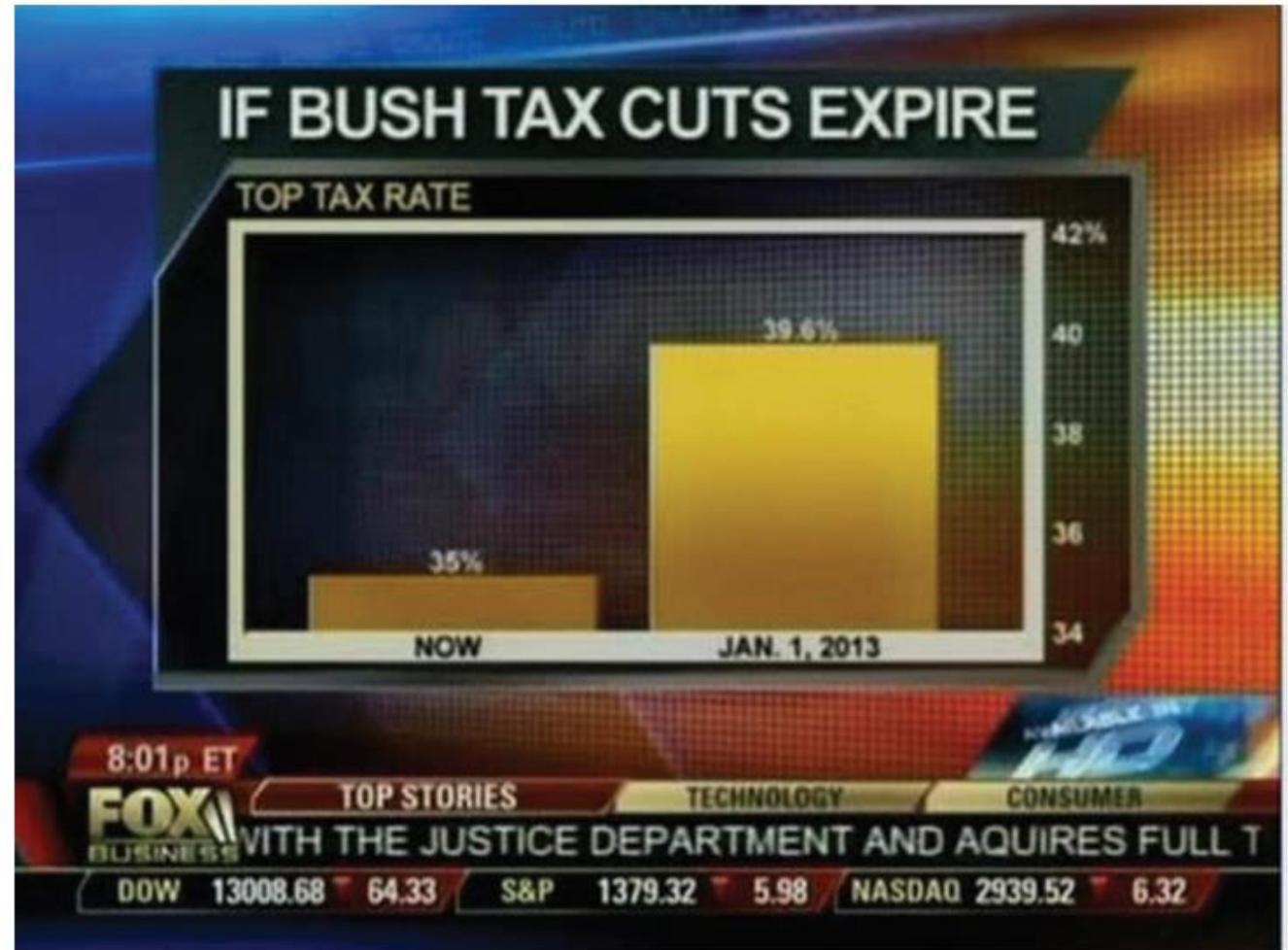


FIGURE 2.12 Fox News bar chart



- Points
- Lines
- Bars
- Area



- For this example, let's imagine we are back in the fall of 2012.
- We are wondering what will happen if the Bush tax cuts expire. On the left-hand side, we have what the top tax rate is currently, 35%, and on the right-hand side what it will be as of January 1, at 39.6%



- Points
- Lines
- Bars
- Area



- Note that the bottom number on the vertical axis (shown at the far right) is not zero, but rather 34.
- This means that the bars, in theory, should continue down through the bottom of the page. In fact, the way this is graphed, the visual increase is 460% (the heights of the bars are  $35 - 34 = 1$  and  $39.6 - 34 = 5.6$ , so  $(5.6 - 1) / 1 = 460\%$ ).

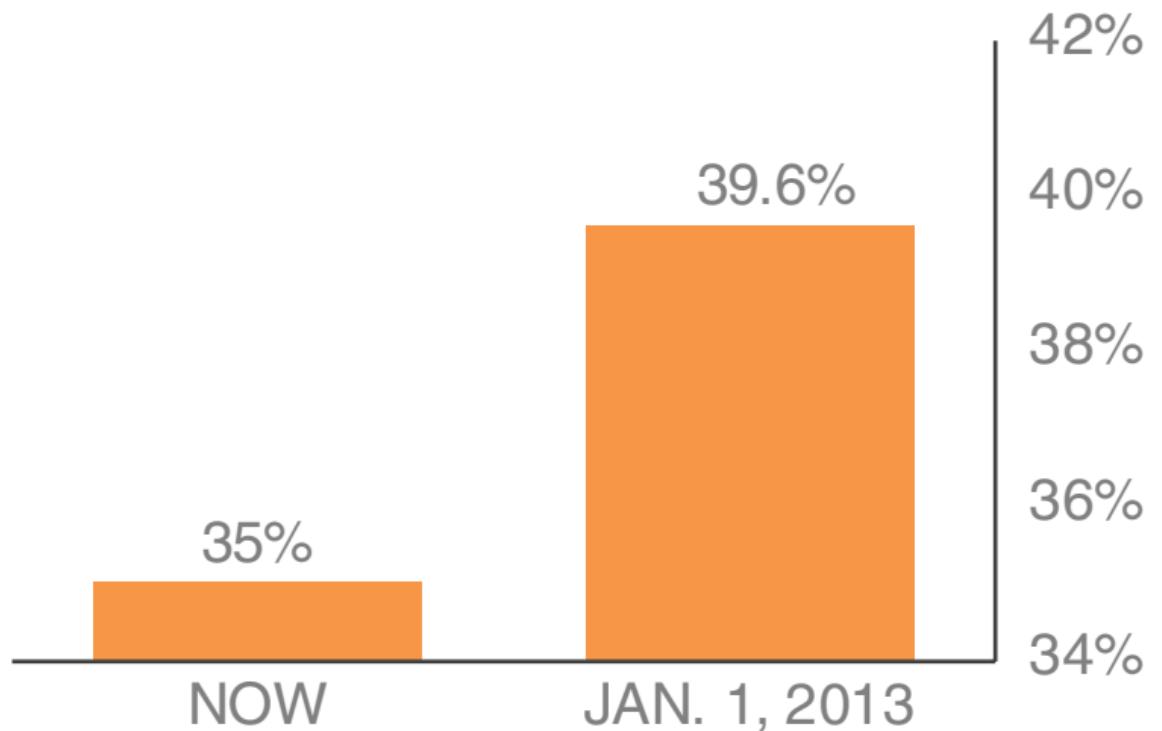


- Points
- Lines
- Bars
- Area



**Non-zero baseline:** as originally graphed

IF BUSH TAX CUTS EXPIRE  
TOP TAX RATE





- Points
- Lines
- Bars
- Area



- If we graph the bars with a zero baseline so that the heights are accurately represented (35 and 39.6), we get an actual visual increase of 13%  $((39.6 - 35) / 35)$ .

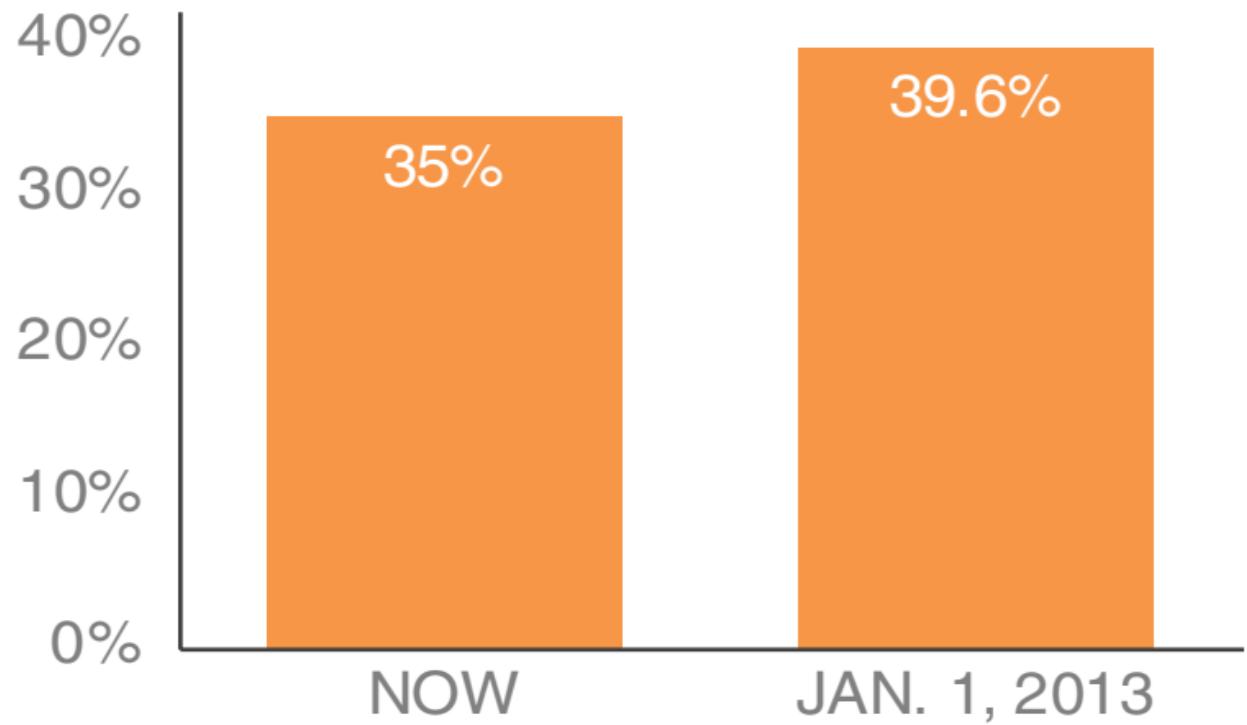


- Points
- Lines
- Bars
- Area



**Zero baseline:** as it should be graphed

IF BUSH TAX CUTS EXPIRE  
TOP TAX RATE



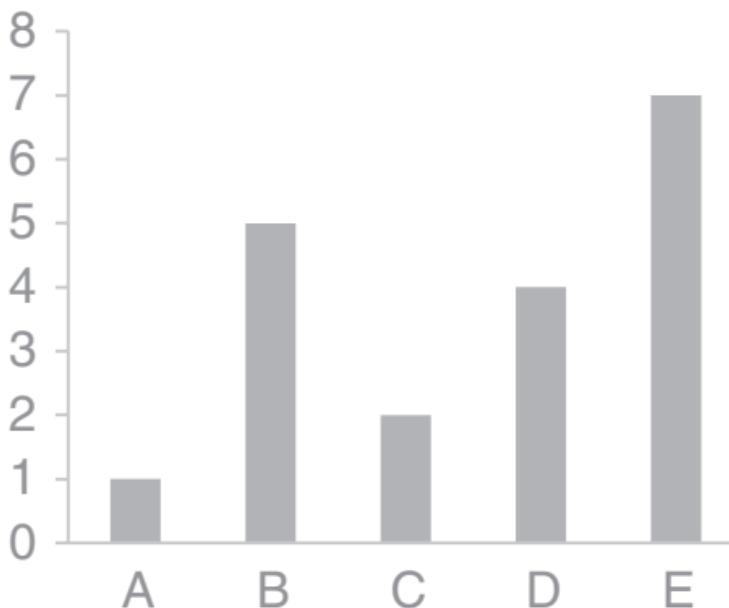


- Points
- Lines
- Bars
- Area



- While we're considering lengths of bars, let's also spend a moment on the width of bars.
- Consider the following “Goldilocks” of bar charts: too thin, too thick, and just right.

Too thin



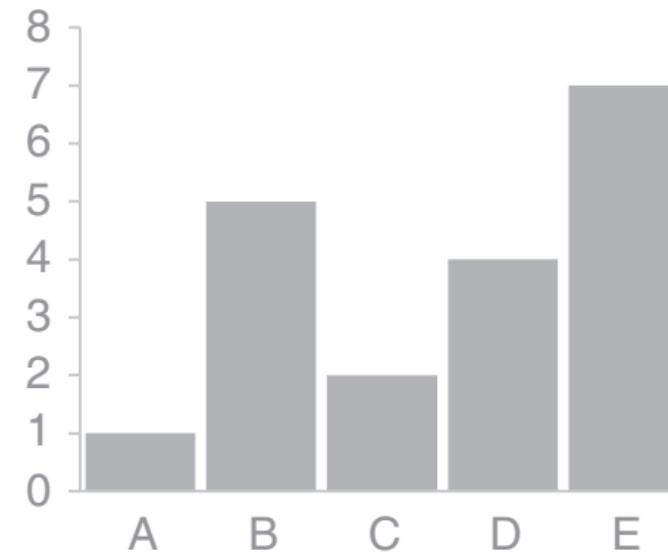


- Points
- Lines
- Bars
- Area



- While we're considering lengths of bars, let's also spend a moment on the width of bars.
- Consider the following “Goldilocks” of bar charts: too thin, too thick, and just right.

Too thick



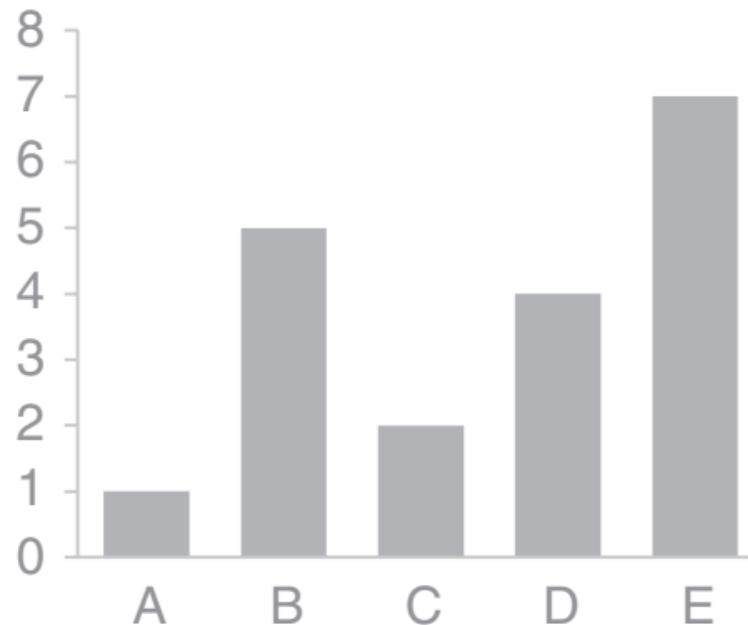


- Points
- Lines
- Bars
- Area



- While we're considering lengths of bars, let's also spend a moment on the width of bars.
- Consider the following “Goldilocks” of bar charts: too thin, too thick, and just right.

Just right





- Points
- Lines
- Bars
- Area



let's look at some different varieties.

- Vertical bar chart
- Stacked vertical bar chart
- Waterfall chart
- Horizontal bar chart
- Stacked horizontal bar chart



- Points
- Lines
- Bars
- Area



## Vertical bar chart

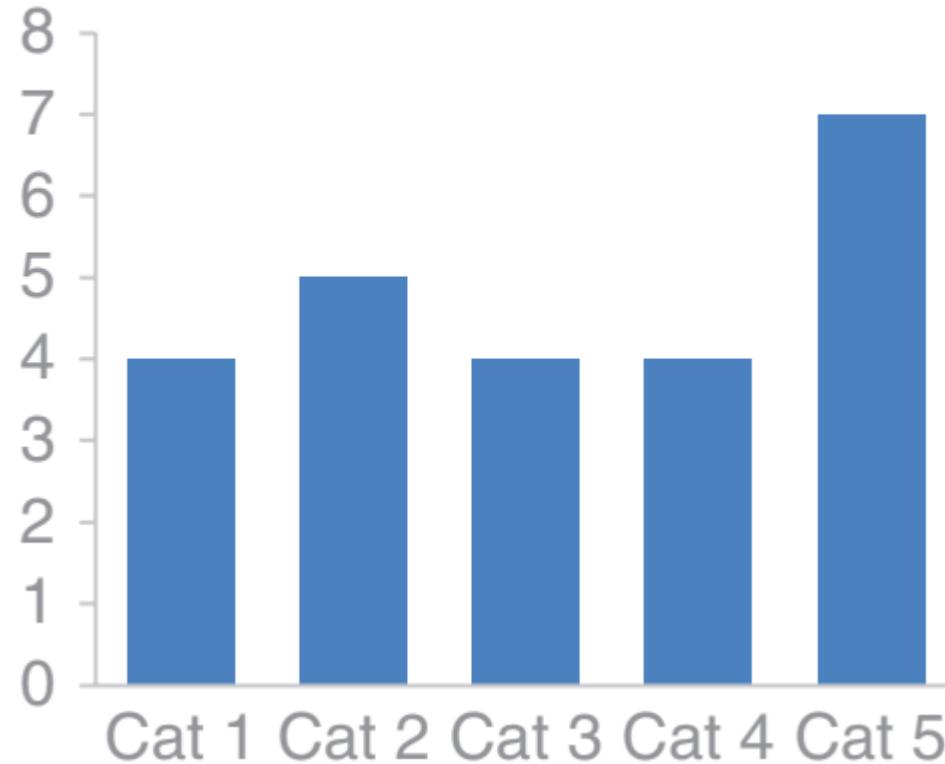
- The plain ordinary bar chart is the vertical bar chart, or column chart.
- Like line graphs, vertical bar charts can be single series, two series, or multiple series. Note that as you add more series of data, it becomes more difficult to focus on one at a time~ and pull~out insight, so use multiple series bar charts with caution.



- Points
- Lines
- Bars
- Area



Single series

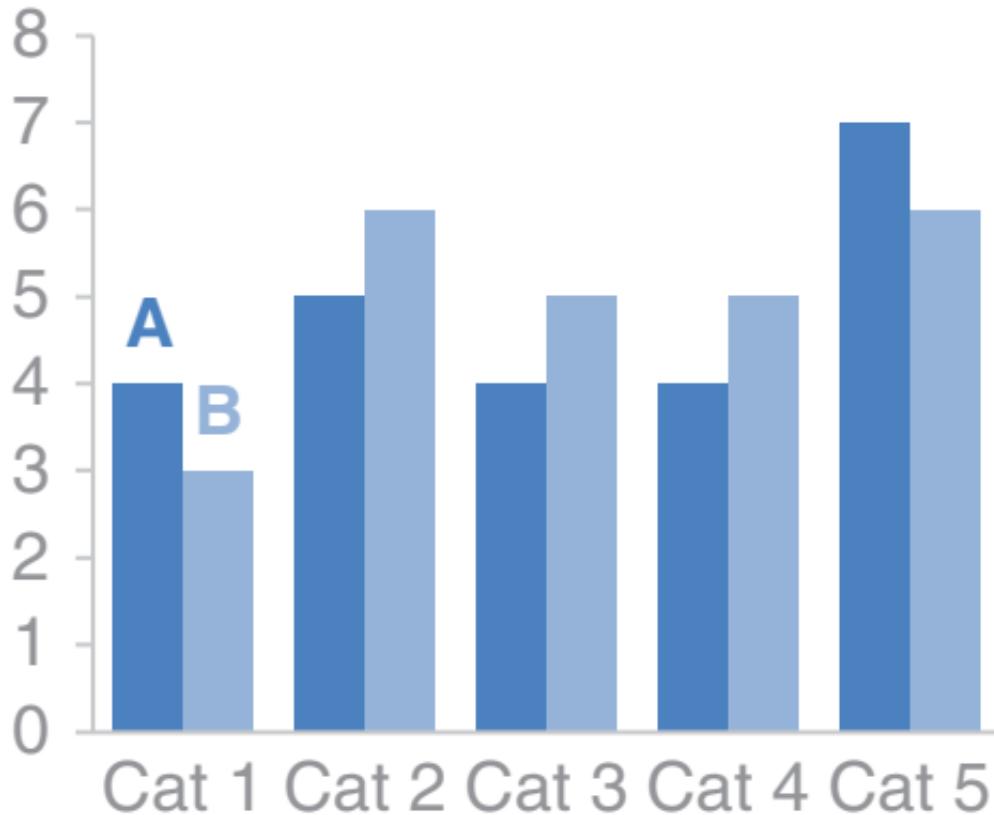




- Points
- Lines
- Bars
- Area



## Two series

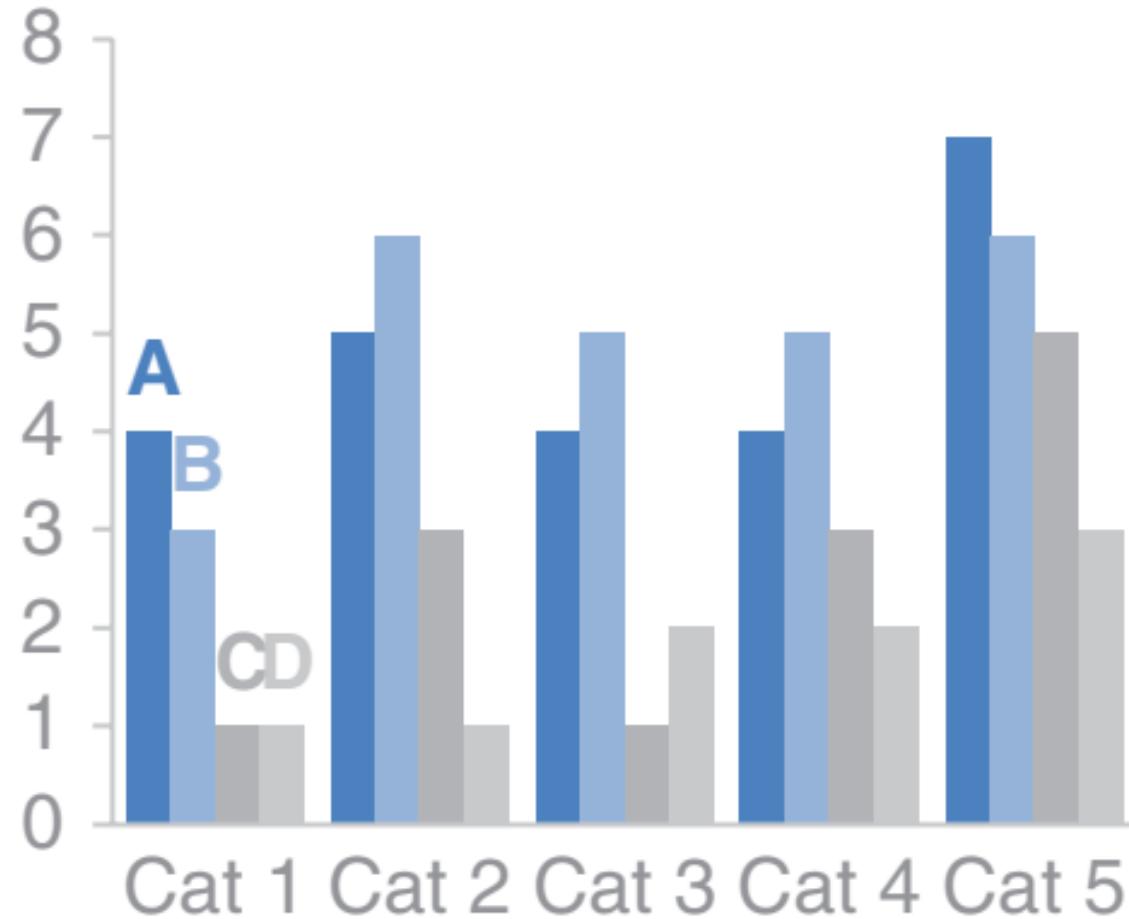




- Points
- Lines
- Bars
- Area



## Multiple series





- Points
- Lines
- Bars
- Area



### Stacked Vertical bar chart

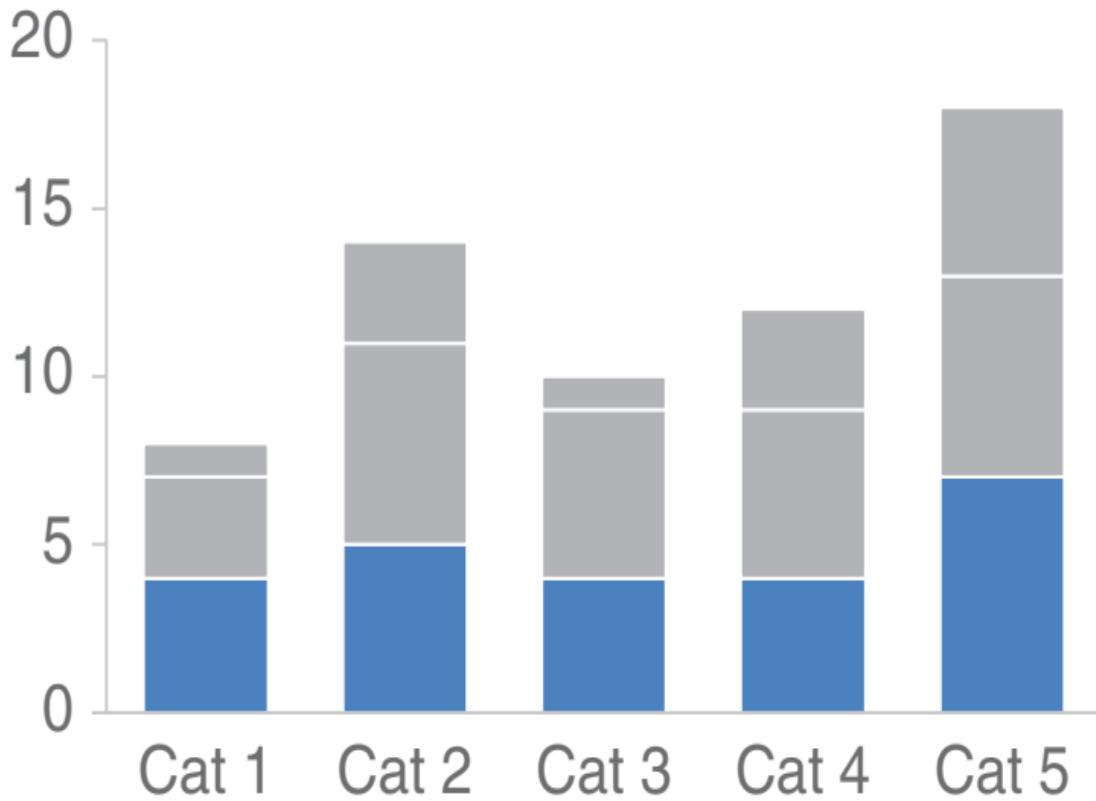
- Use cases for stacked vertical bar charts are more limited. They are meant to allow you to compare totals across categories and see the subcomponent pieces within a given category.



- Points
- Lines
- Bars
- Area



Comparing **these** is easy

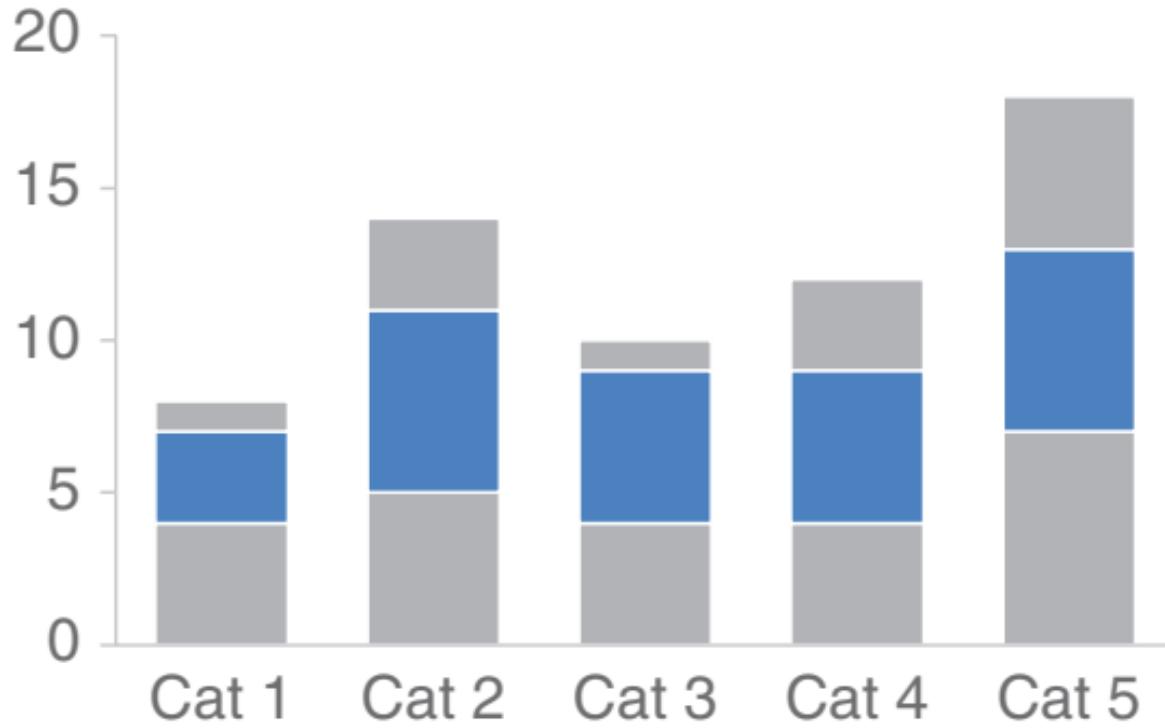




- Points
- Lines
- Bars
- Area



Comparing **these** is hard





- Points
- Lines
- Bars
- Area



## Waterfall chart

- The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.



- Points
- Lines
- Bars
- Area



### 2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.

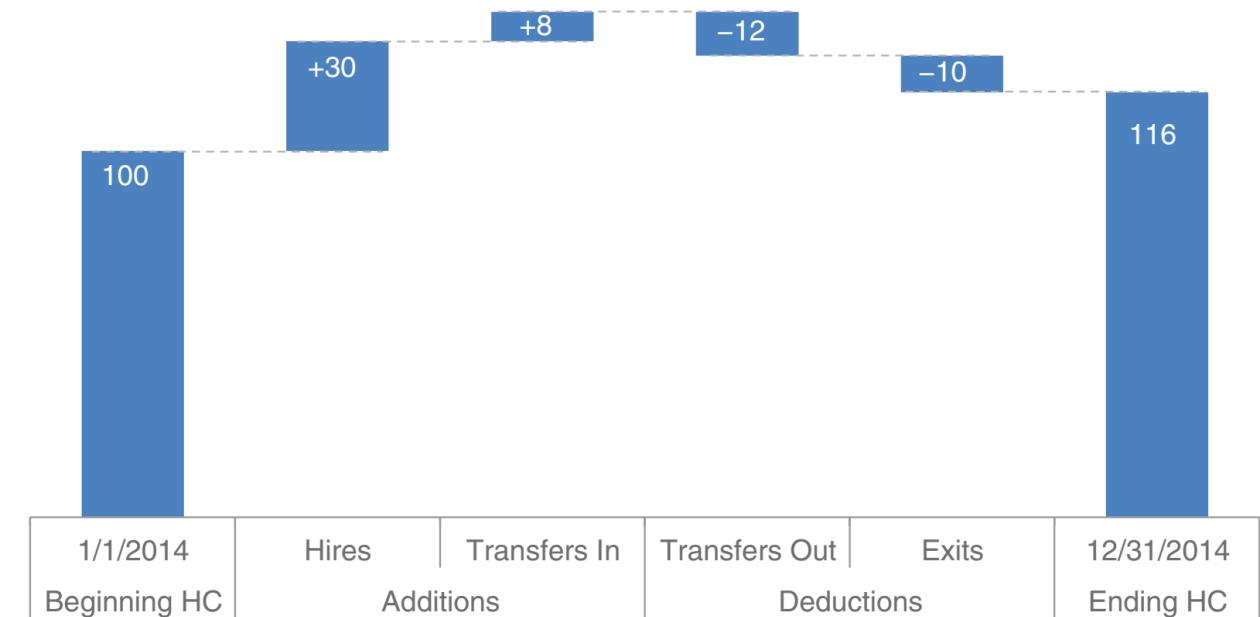
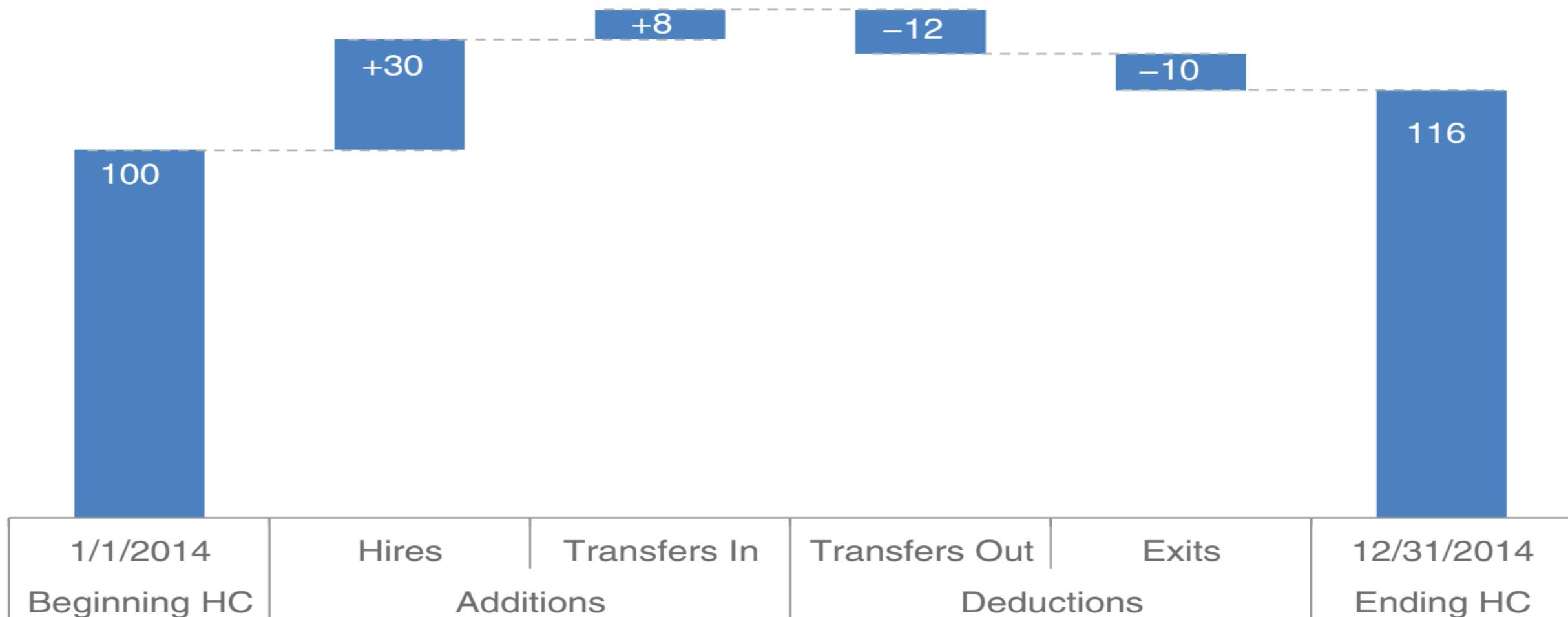


FIGURE 2.17 Waterfall chart

## 2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.



**FIGURE 2.17** Waterfall chart



- Points
- Lines
- Bars
- Area



### 2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.

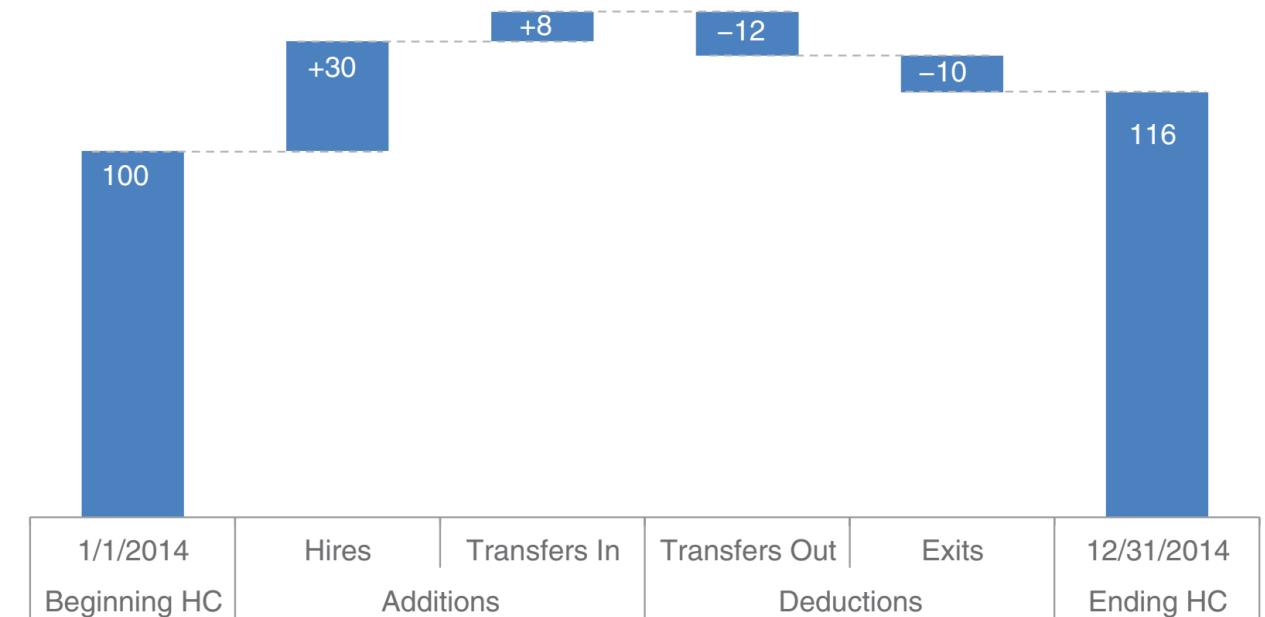


FIGURE 2.17 Waterfall chart



- Points
- Lines
- Bars
- Area



## Horizontal bar chart

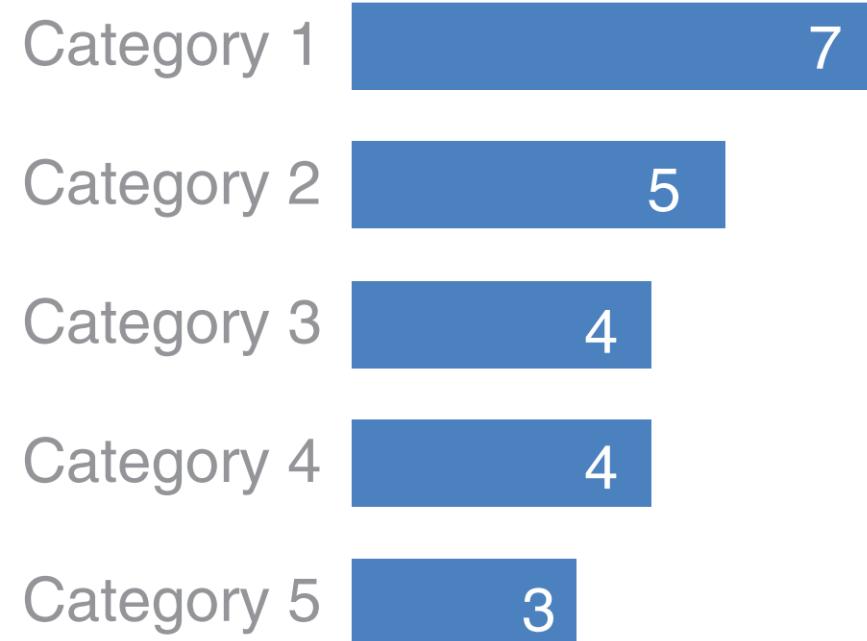
- The horizontal bar chart is especially useful if your category names are long, as the text is written from left to right, as most audiences read, making your graph legible for your audience.
- Like the vertical bar chart, the horizontal bar chart can be single series, two series, or multiple series



- Points
- Lines
- Bars
- Area



## Single series

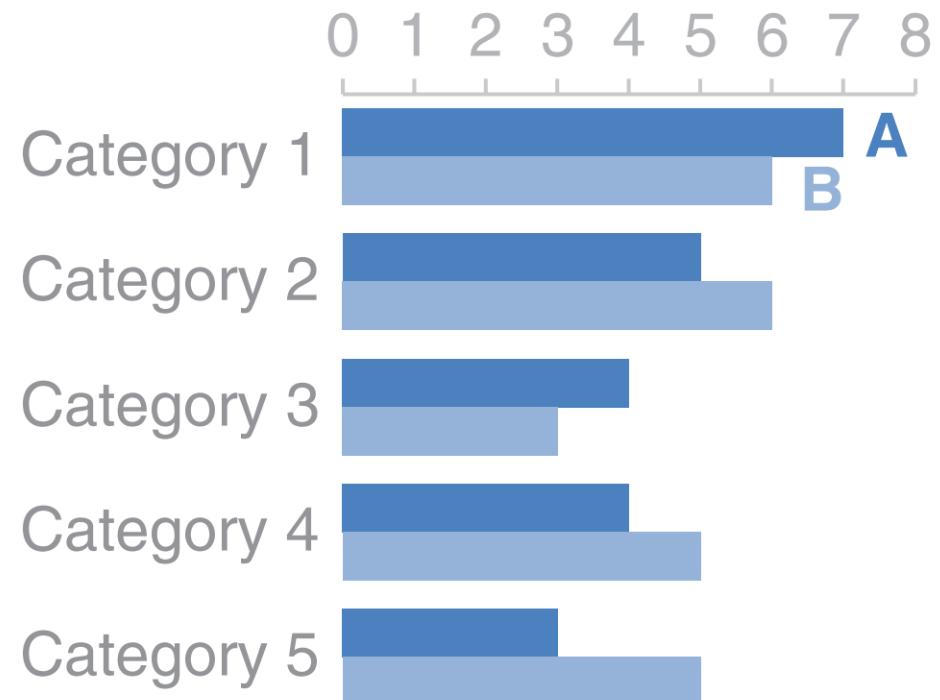




- Points
- Lines
- Bars
- Area



## Two series

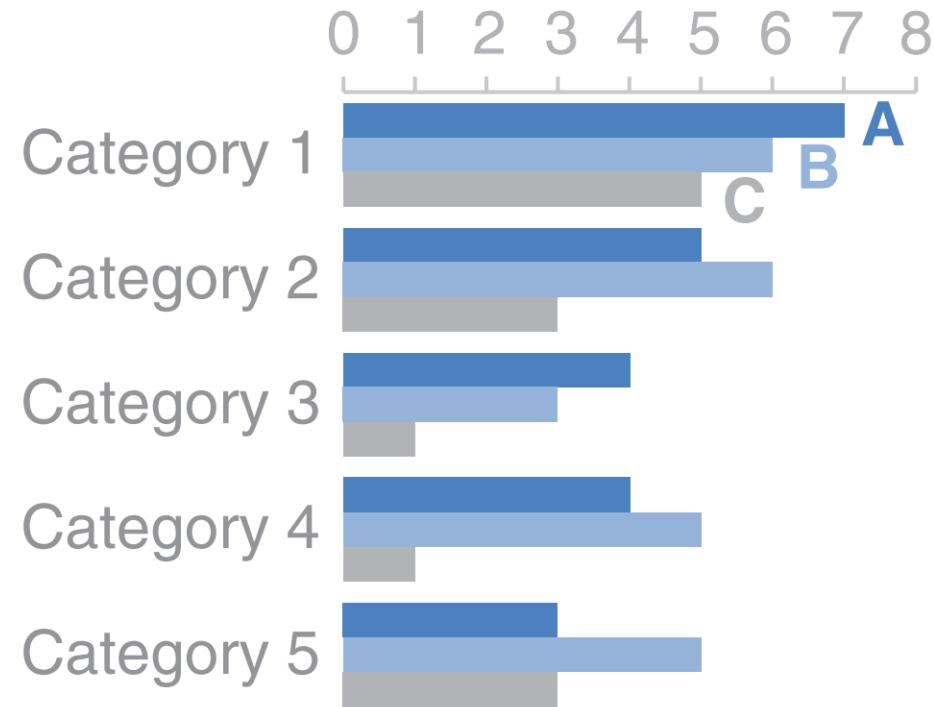




- Points
- Lines
- Bars
- Area



## Multiple series





- Points
- Lines
- Bars
- Area



### Stacked Horizontal bar chart

- Like the stacked vertical bar chart, stacked horizontal bar charts can be used to show the totals across different categories but also give a sense of the subcomponent pieces.



- Points
- Lines
- Bars
- Area



## Survey results

Strongly Disagree | Disagree | Neutral | **Agree** | Strongly Agree

Percent of total

0% 20% 40% 60% 80% 100%



FIGURE 2.19 100% stacked horizontal bar chart



- Points
- Lines
- Bars
- Area



### Area:

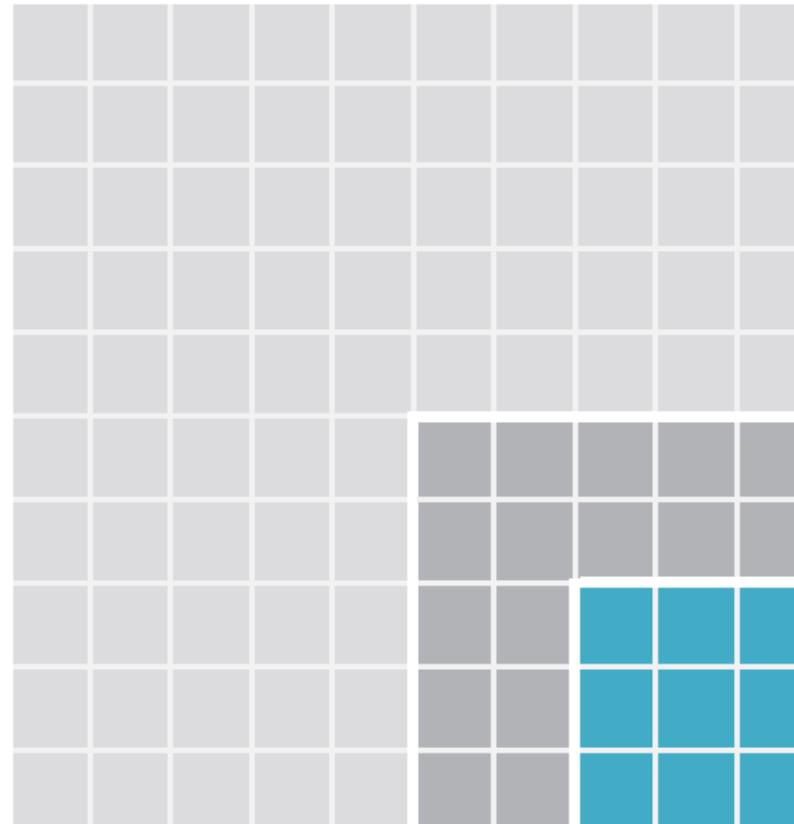
- Avoid most area graphs. Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read.



- Points
- Lines
- Bars
- Area



## Interview breakdown



Out of every 100 phone screens...

we bring **25 candidates onsite** for interviews...

and **extend 9 offers.**

**FIGURE 2.20** Square area graph



## To be avoided:

- There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary y-axes.

Supplier Market Share

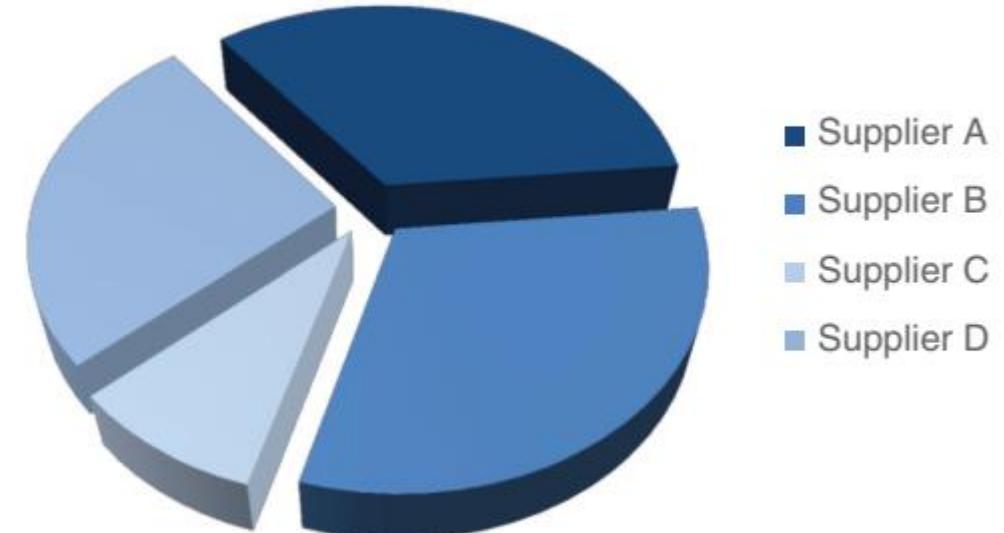


FIGURE 2.21 Pie chart



## To be avoided:

- There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary y-axes.

The donut chart

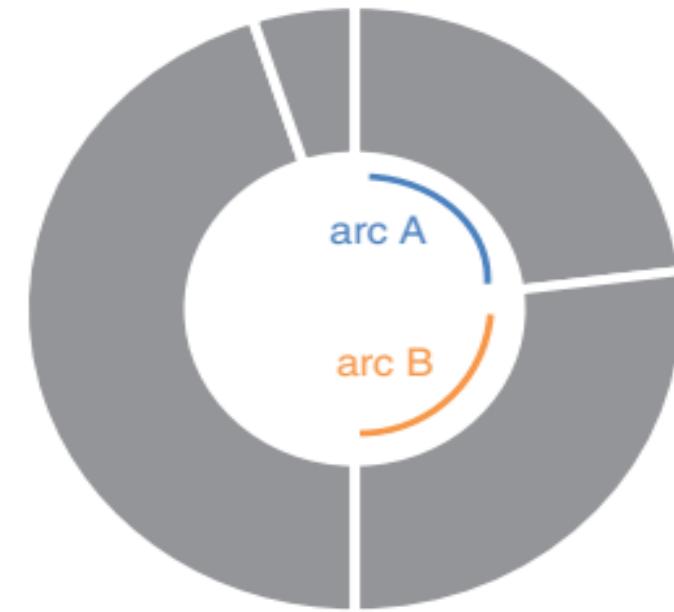


FIGURE 2.24 Donut chart



## To be avoided:

- There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary y-axes.

Number of issues

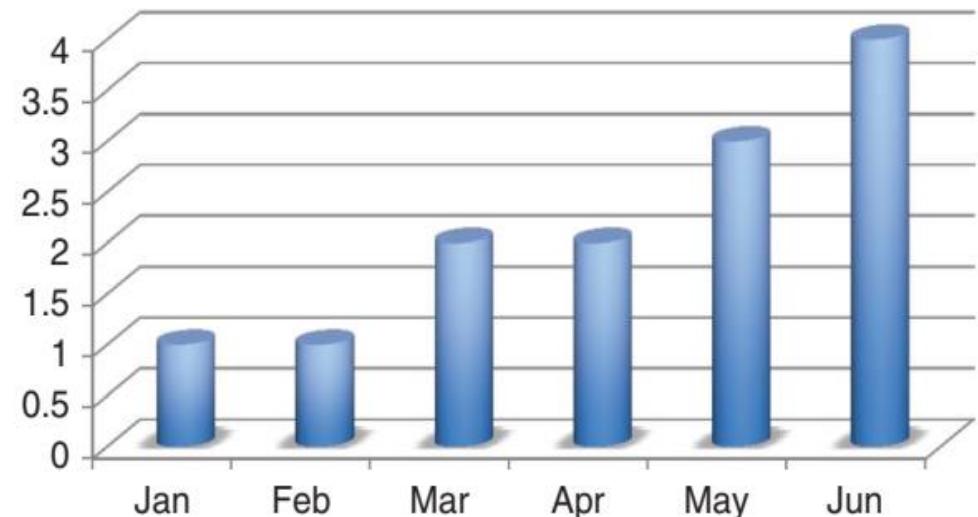


FIGURE 2.25 3D column chart



### To be avoided:

- There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary y-axes.

Secondary y-axis

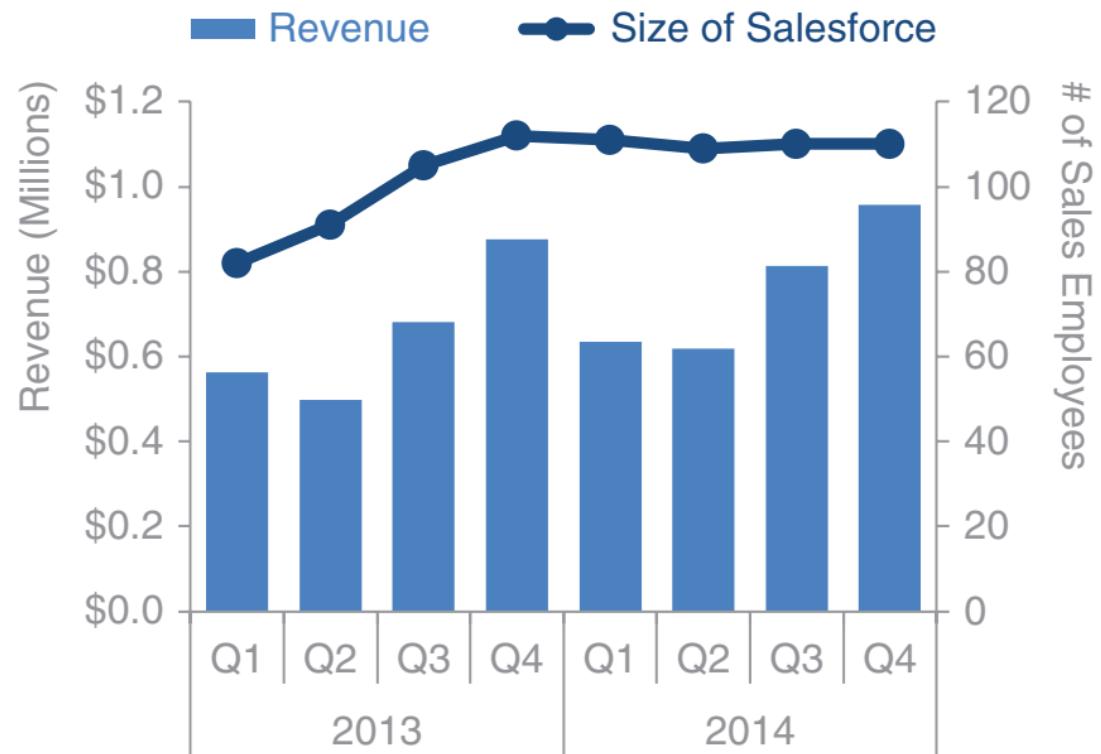


FIGURE 2.26 Secondary y-axis



# Contents

1

Introduction

2

The Importance of Context

3

Choosing the Effective Visual



# DATA VISUALIZATION



# UNIT - II



# Contents

1

Clutter is your enemy

2

Focus your audience's attention

3

Lessons in Storytelling

# How you'll learn to tell stories with data



## STEP 1

### CONTEXT

Understand the Context



## STEP 2

### DISPLAY

Choose an appropriate visual display



## STEP 3

### CLUTTER

Eliminate Clutter



## STEP 4

### ATTENTION

Focus attention where you want it



## STEP 5

### THINKING

Think like a designer



## STEP 6

### TELL A STORY

Tell a story to audience





# Clutter is your enemy

- Picture a blank page or a blank screen: every single element you add to that page or screen takes up cognitive load on the part of your audience—in other words, takes their brain power to process.

## Cognitive load:

- Perhaps you were sitting in a conference room as the person leading the meeting was flipping through their projected slides and they paused on one that looked overwhelmingly busy and complicated.



# Clutter is your enemy

- Yikes, did you say “ugh” out loud, or was that just in your head?
- Maybe you were reading through a report or the newspaper, and a graph caught your eye just long enough for you to think, “ this looks interesting, but I have no idea what I’m meant to get out of it”

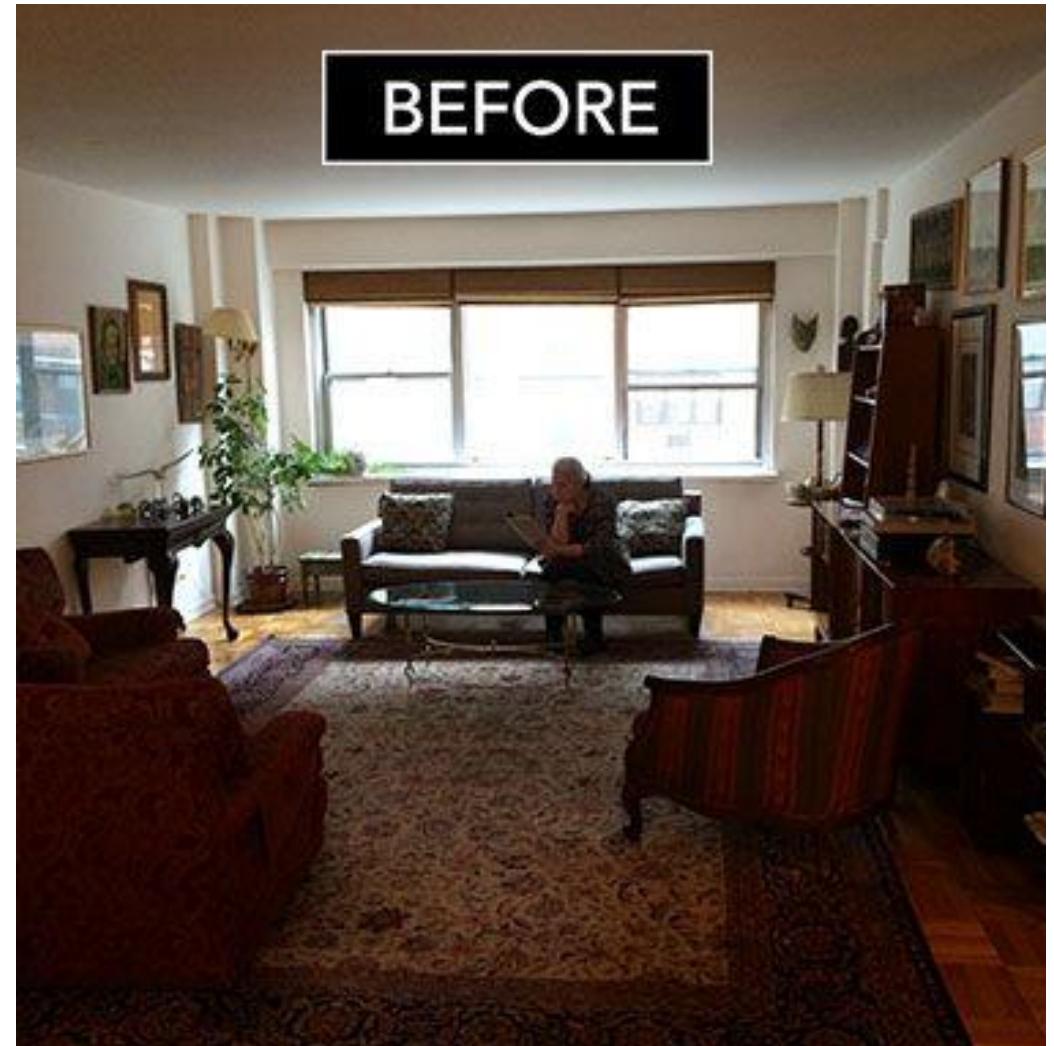




# Clutter is your enemy

## Clutter:

- One culprit that can contribute to excessive or extraneous cognitive load is something refer to simply as clutter.
- These are visual elements that take up space but don't increase understanding.





# Clutter is your enemy

## Clutter:

- There is a simple reason we should aim to reduce clutter: because it makes our visuals appear more complicated than necessary.

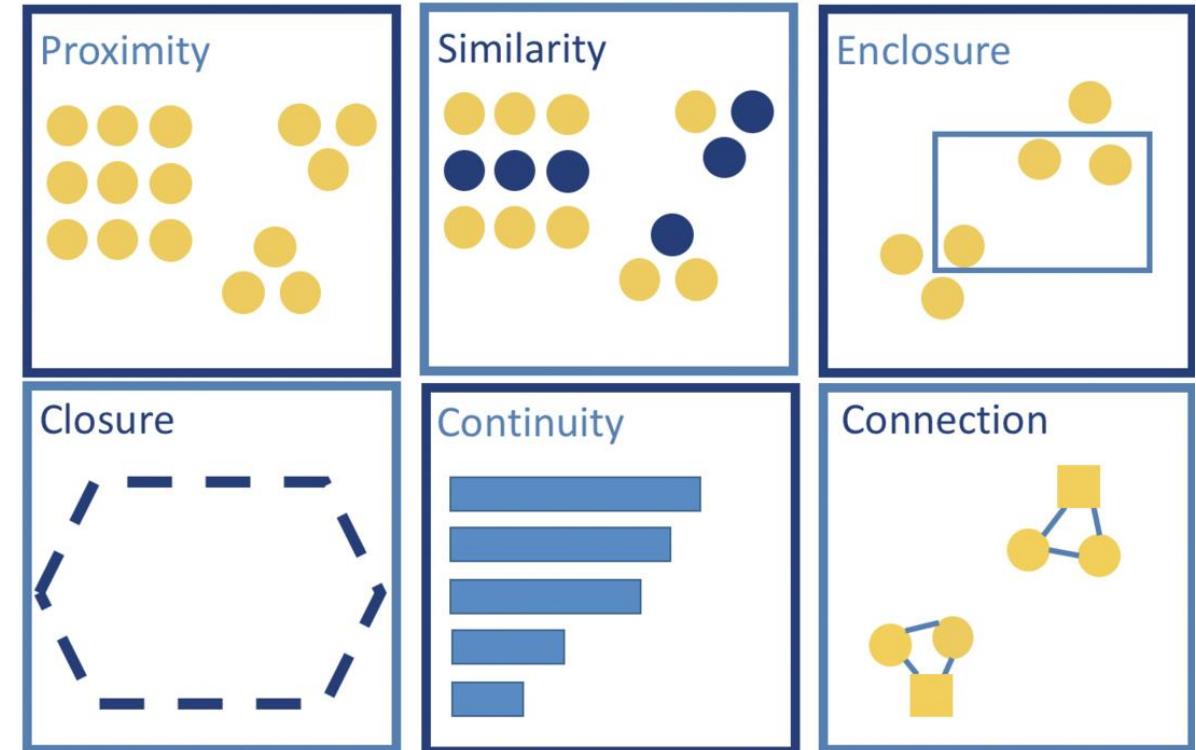




# Clutter is your enemy

## Gestalt principles of visual perception:

- The Gestalt School of Psychology set out in the early 1900s to understand how individuals perceive order in the world around them.
- **Six principles** : proximity, similarity, enclosure, closure, continuity, and connection.





Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- We tend to think of objects that are physically close together as belonging to part of a group.





# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

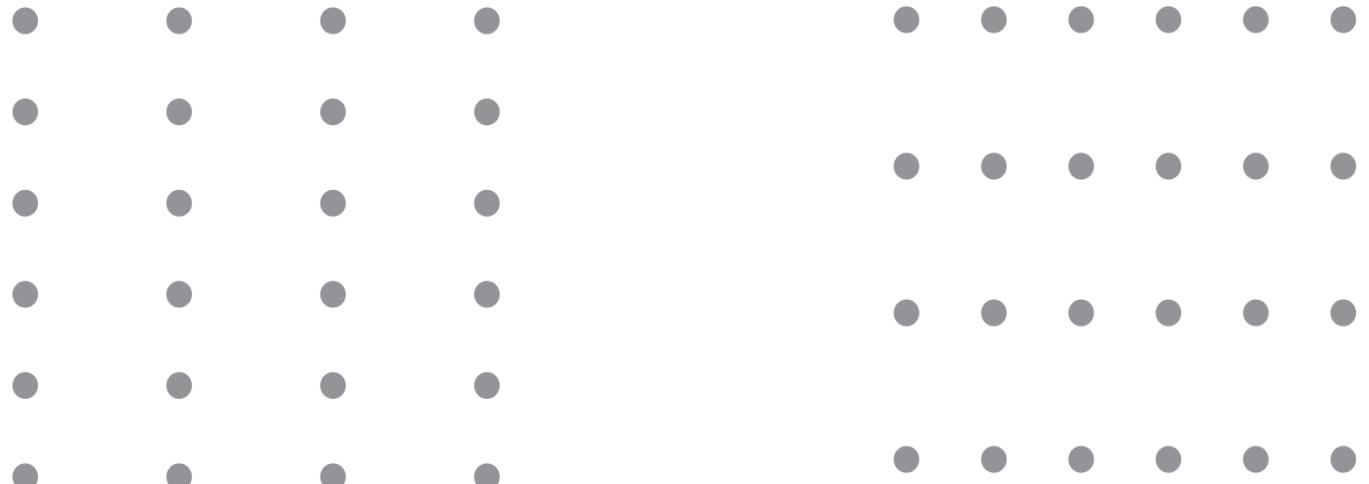
- We can leverage this way that people see in table design.
- In Figure simply by virtue of differentiating the spacing between the dots, your eyes are drawn either down the columns in the first case or across the rows in the second case.



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection





# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- **Similarity**
- Enclosure
- Closure
- Continuity
- Connection

- Objects that are of similar color, shape, size, or orientation are perceived as related or belonging to part of a group.
- In Figure you naturally associate the blue circles together on the left or the grey squares together on the right.



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- **Similarity**
- Enclosure
- Closure
- Continuity
- Connection



**FIGURE 3.3** Gestalt principle of similarity

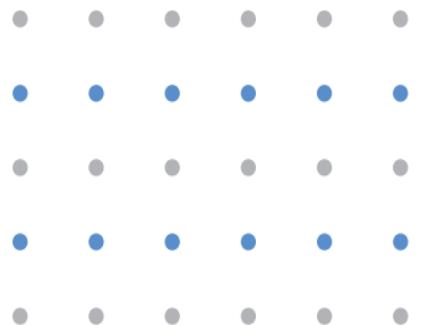


Gestalt principles of visual perception:

- Proximity
- **Similarity**
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- In Figure 3.4, the similarity of color is a cue for our eyes to read across the rows.
- This eliminates the need for additional elements such as borders to help direct our attention



**FIGURE 3.4** You see rows due to similarity of color



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

- We think of objects that are physically enclosed together as belonging to part of a group.
- It doesn't take a very strong enclosure to do this: light background shading is often enough, as demonstrated in Figure 3.5



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- **Enclosure**
- Closure
- Continuity
- Connection



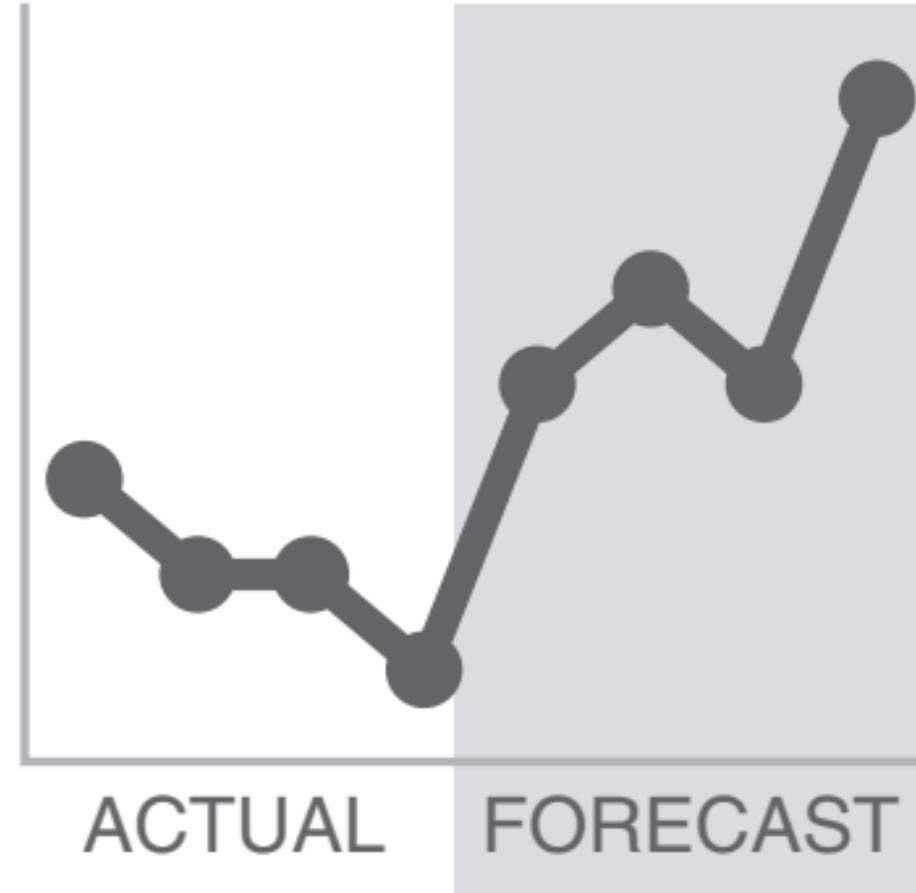
**FIGURE 3.5** Gestalt principle of enclosure



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection





# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

- The closure concept says that people like things to be simple and to fit in the constructs that are already in our heads.
- Because of this, people tend to perceive a set of individual elements as a single



Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- For example, the elements in Figure 3.7 will tend to be perceived as a circle first and only after that as individual elements.



FIGURE 3.7 Gestalt principle of closure



Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- It is common for graphing applications to have default settings that include elements like chart borders and background shading.



**FIGURE 3.8** The graph still appears complete without the border and background shading



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection
- The principle of continuity is like closure: when looking at objects, our eyes seek the smoothest path and naturally create continuity in what we see even where it may not explicitly exist.



Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- Figure 3.9, if I take the objects (1) and pull them apart, most people will expect to see what is shown next (2), whereas it could as easily be what is shown after that (3).

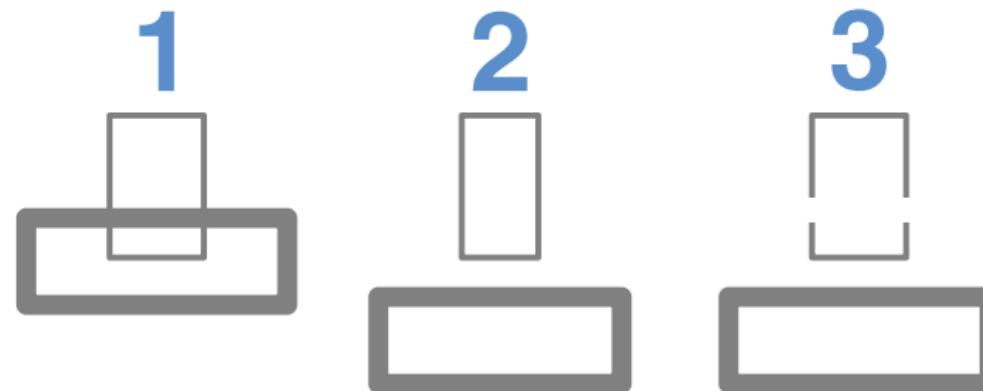


FIGURE 3.9 Gestalt principle of continuity

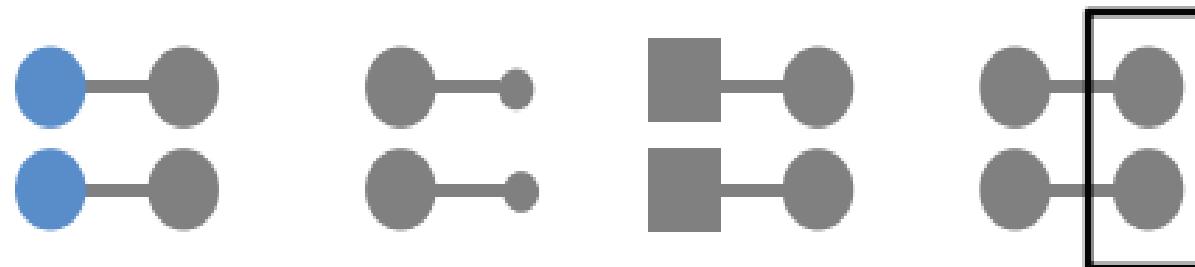


Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

# Clutter is your enemy

- We tend to think of objects that are physically connected as part of a group.
- The connective property typically has a stronger associative value than similar color, size, or shape.



**FIGURE 3.11** Gestalt principle of connection



# Clutter is your enemy

Gestalt principles of visual perception:

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity
- Connection

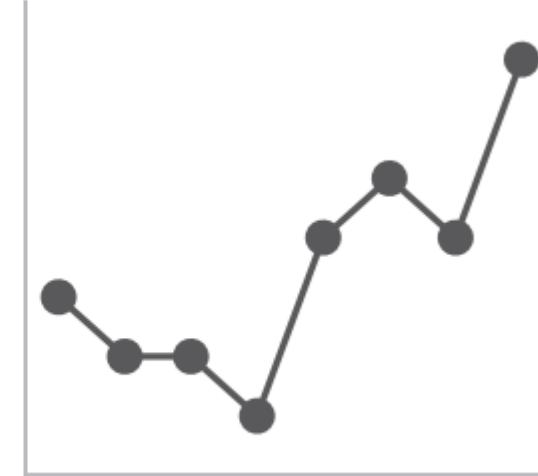
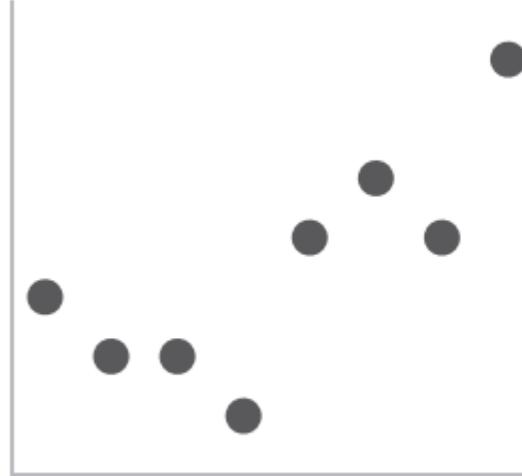


FIGURE 3.12 Lines connect the dots



# Clutter is your enemy

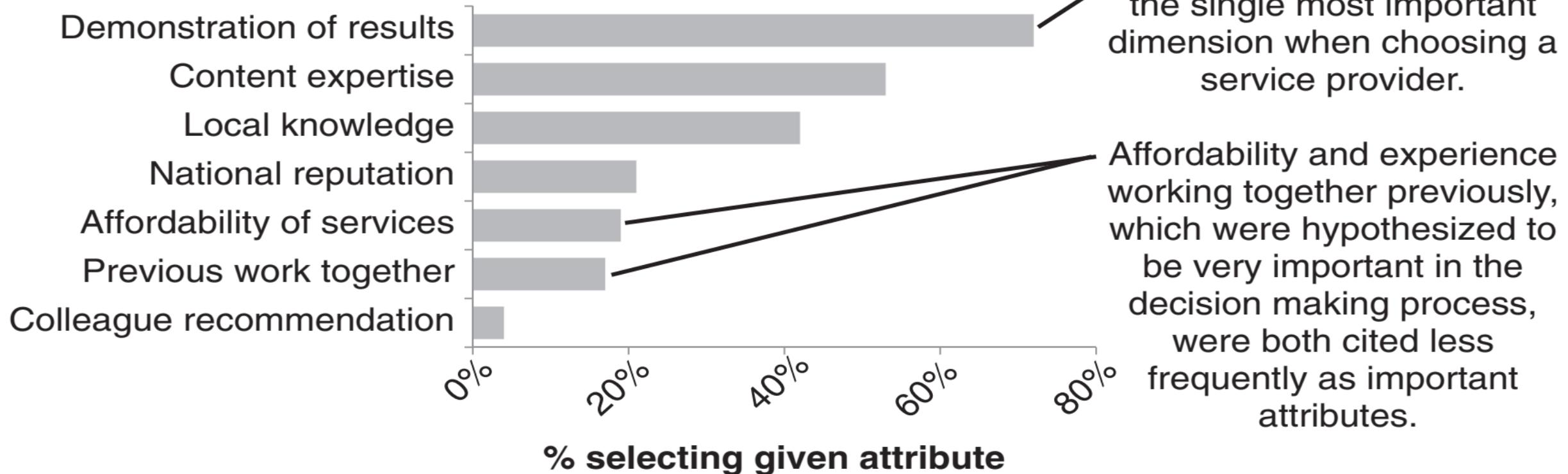
let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- When design is thoughtful, it fades into the background so that your audience doesn't even notice it.
- Let's look at an example to understand the impact visual order—and lack thereof—can have on our visual communications.

## Demonstrating effectiveness is most important consideration when selecting a provider

In general, what attributes are the most important to you in selecting a service provider?  
*(Choose up to 3)*



Survey shows that demonstration of results is the single most important dimension when choosing a service provider.

Affordability and experience working together previously, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

**FIGURE 3.13** Summary of survey feedback



# Clutter is your enemy

let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- Take a moment to study Figure 3.13, which summarizes survey feedback about factors considered by nonprofits in vendor selection.

## Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?

(Choose up to 3)



Demonstration of results

Survey shows that **demonstration of results** is the single most important dimension when choosing a service provider.

Content expertise

Local knowledge

National reputation

Affordability of services

Previous work together

Colleague recommendation

**Affordability** and **experience working together previously**, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

Data source: xyz; includes N number of survey respondents.  
Note that respondents were able to choose up to 3 options.

**FIGURE 3.14** Revamped summary of survey feedback



# Clutter is your enemy

let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- We can improve this visual markedly by making some relatively minor changes.
- Look at Figure 3.14. The content is the same; only the placement and formatting of elements have been modified.



# Clutter is your enemy

let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- The lack of clear contrast, on the other hand, can be a form of visual clutter.
- Ex: Imagine you work for a U.S. retailer and want to understand how your customers feel about various dimensions of their shopping experience in your store compared to your competitors.



# Clutter is your enemy

let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- Figure 3.15 shows the weighted performance index across categories for your company and five competitors.



FIGURE 3.15 Original graph

## Weighted Performance Index

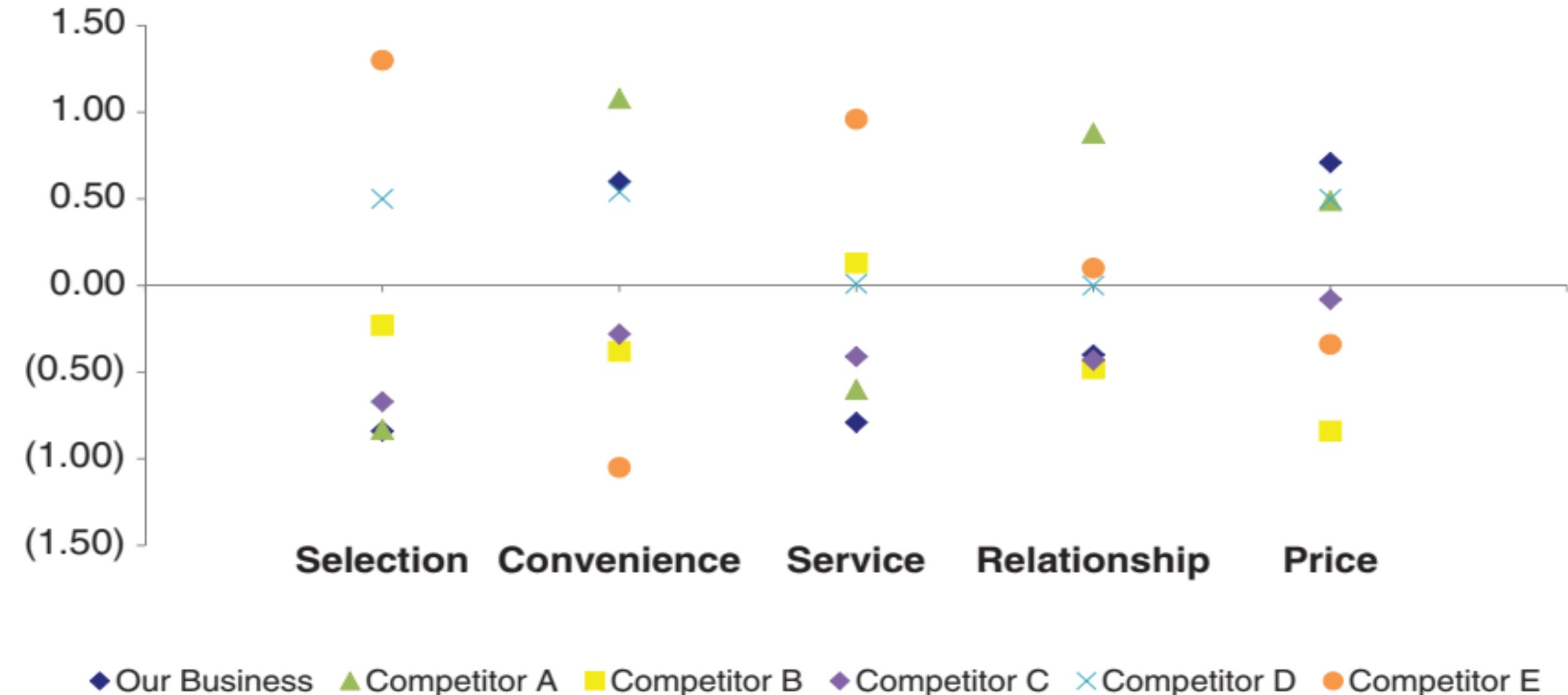


FIGURE 3.15 Original graph



# Clutter is your enemy

let's shift our focus to a couple of other types of visual clutter.

- Lack of visual order
- Non-strategic use of contrast

- Figure 3.15 shows the weighted performance index across categories for your company and five competitors.



FIGURE 3.15 Original graph

# Performance overview

## ■ Our business

- Competitor A
- Competitor B
- Competitor C
- Competitor D
- Competitor E



FIGURE 3.16 Revamped graph, using contrast strategically



# Clutter is your enemy

## Decluttering: step-by-step:

- Ex: Imagine that you manage an information technology (IT) team. Your team receives tickets, or technical issues, from employees.
- In the past year, you've had a couple of people leave and decided at the time not to replace them.

- You have heard a rumbling of complaints from the remaining employees about having to “pick up the slack.”
- You’ve just been asked about your hiring needs for the coming year and are wondering if you should hire a couple more people.



# Clutter is your enemy

## Decluttering: step-by-step:

- Ex: Imagine that you manage an information technology (IT) team. Your team receives tickets, or technical issues, from employees.
- In the past year, you've had a couple of people leave and decided at the time not to replace them.

- You have heard a rumbling of complaints from the remaining employees about having to “pick up the slack.”
- You’ve just been asked about your hiring needs for the coming year and are wondering if you should hire a couple more people.



# Clutter is your enemy

## Decluttering: step-by-step:

- Ex: Imagine that you manage an information technology (IT) team. Your team receives tickets, or technical issues, from employees.
- In the past year, you've had a couple of people leave and decided at the time not to replace them.

- You have heard a rumbling of complaints from the remaining employees about having to “pick up the slack.”
- You’ve just been asked about your hiring needs for the coming year and are wondering if you should hire a couple more people.

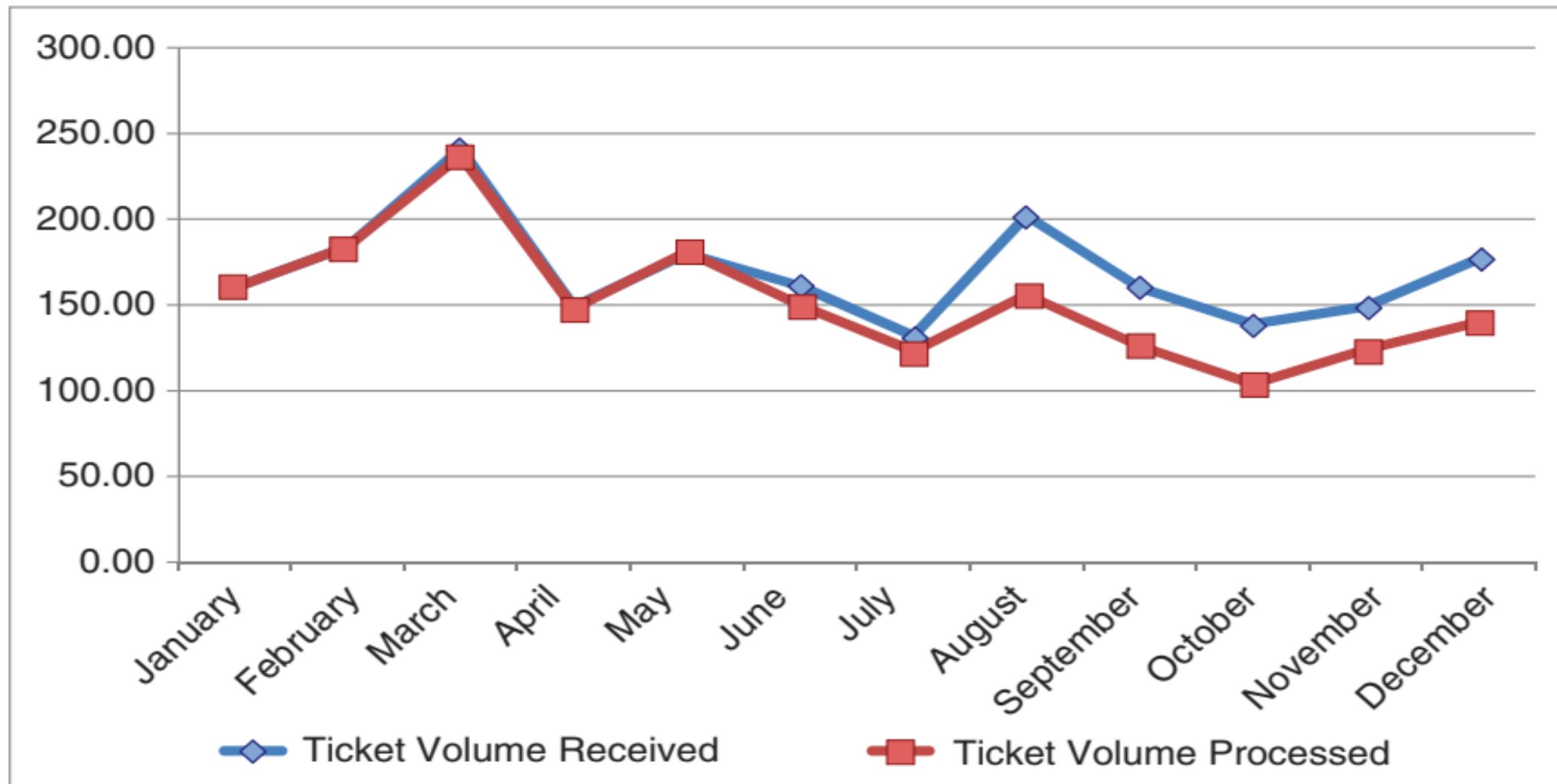


FIGURE 3.17 Original graph



# Clutter is your enemy

Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color

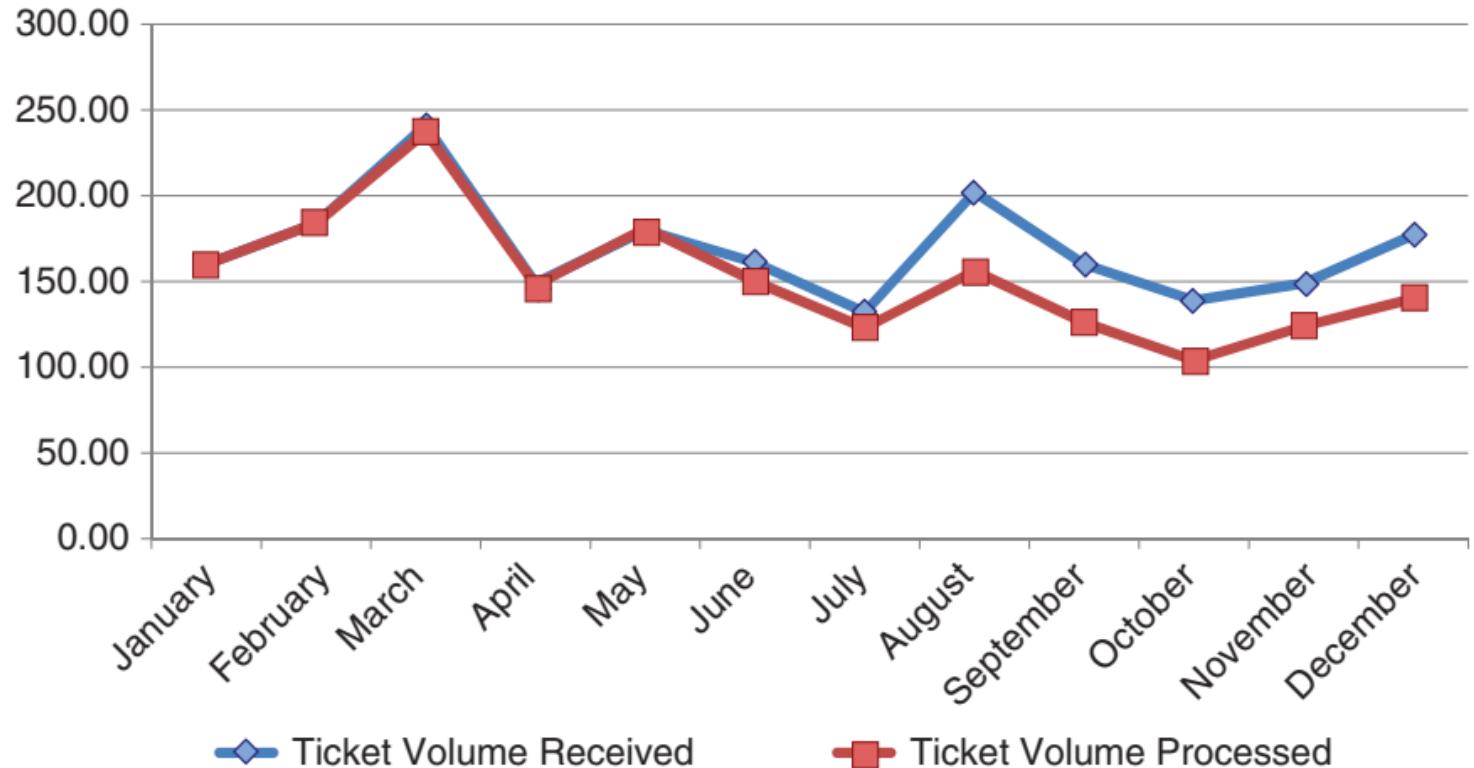


FIGURE 3.18 Remove chart border



# Clutter is your enemy

## Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color

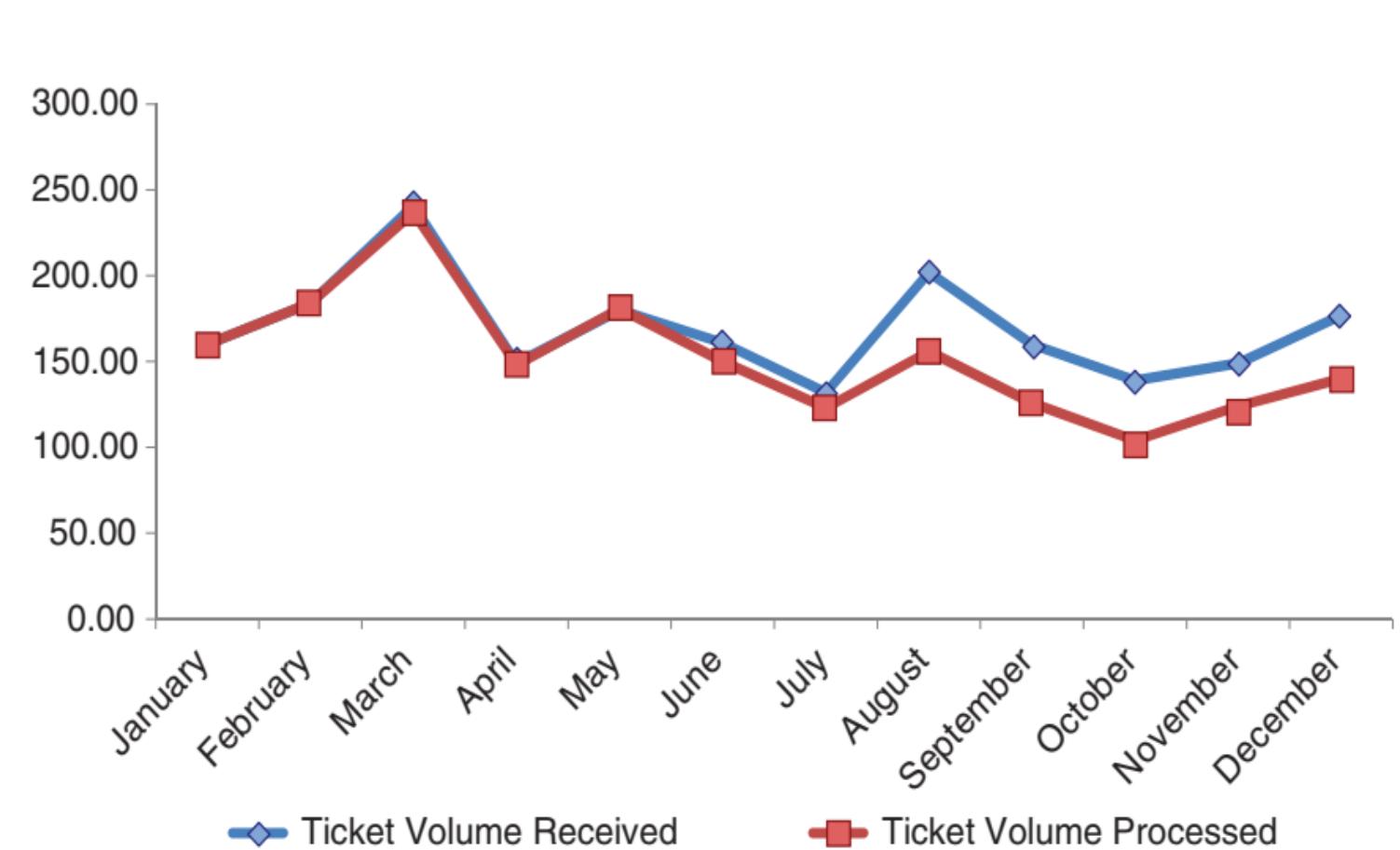


FIGURE 3.19 Remove gridlines



# Clutter is your enemy

## Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color

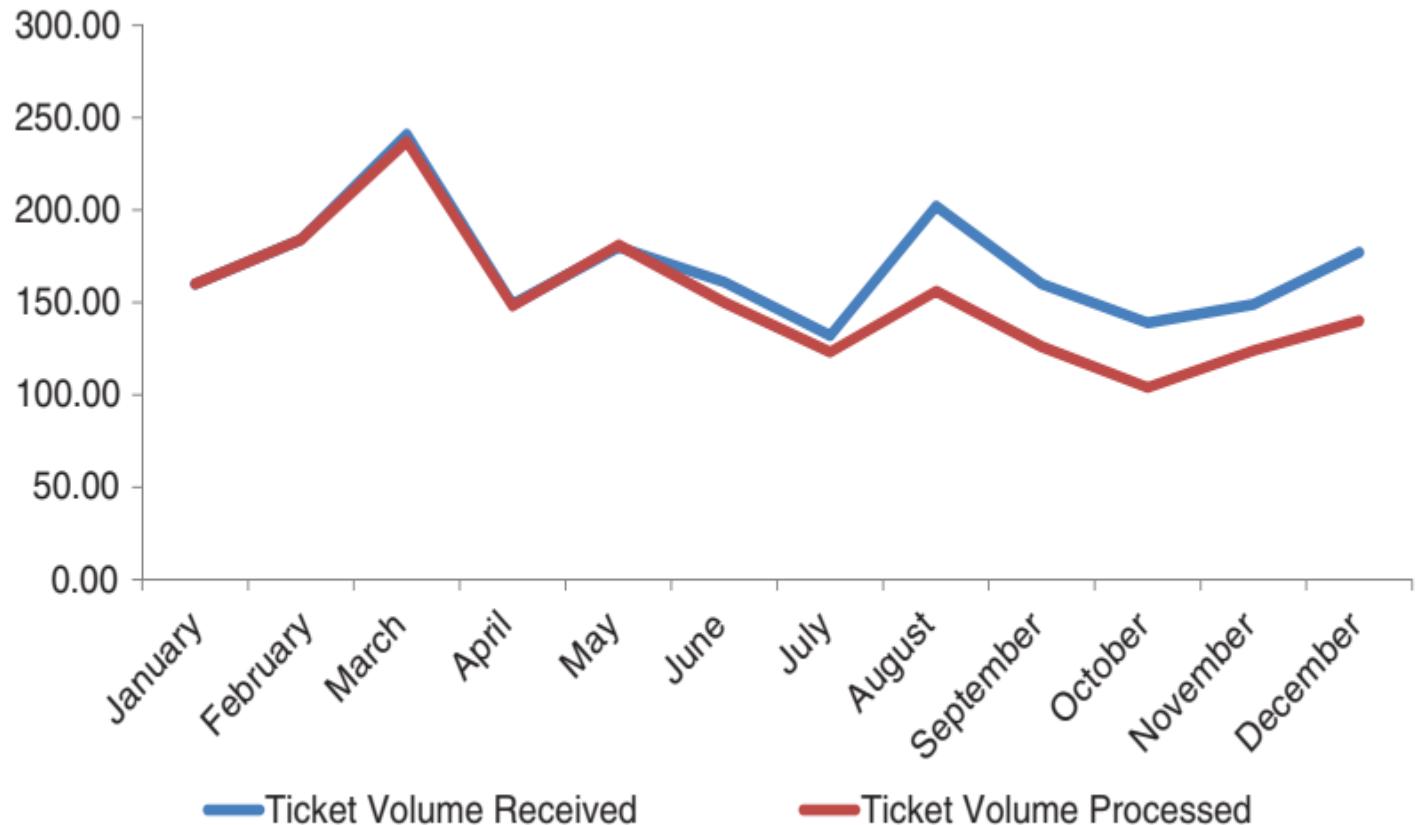


FIGURE 3.20 Remove data markers



# Clutter is your enemy

Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. **Clean up axis labels**
5. Label data directly
6. Leverage consistent color

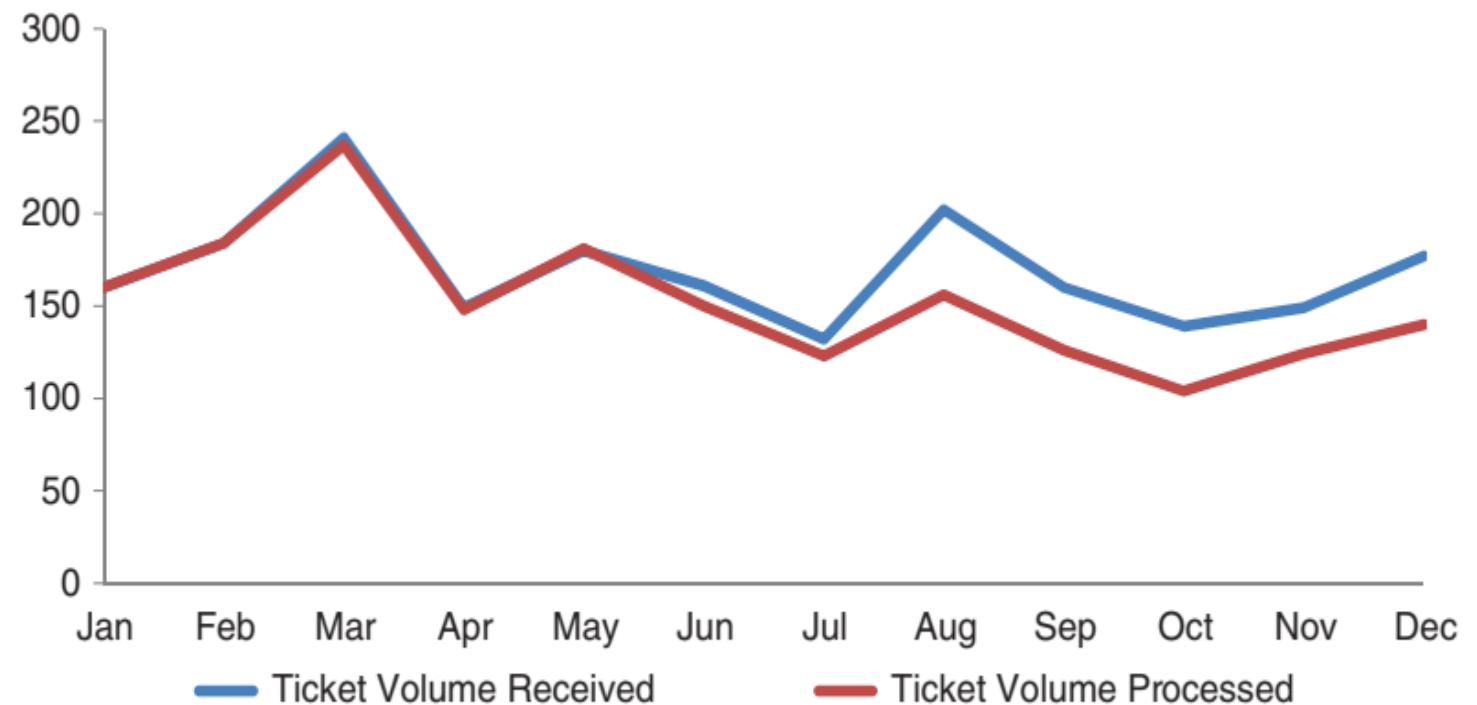


FIGURE 3.21 Clean up axis labels



# Clutter is your enemy

## Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color

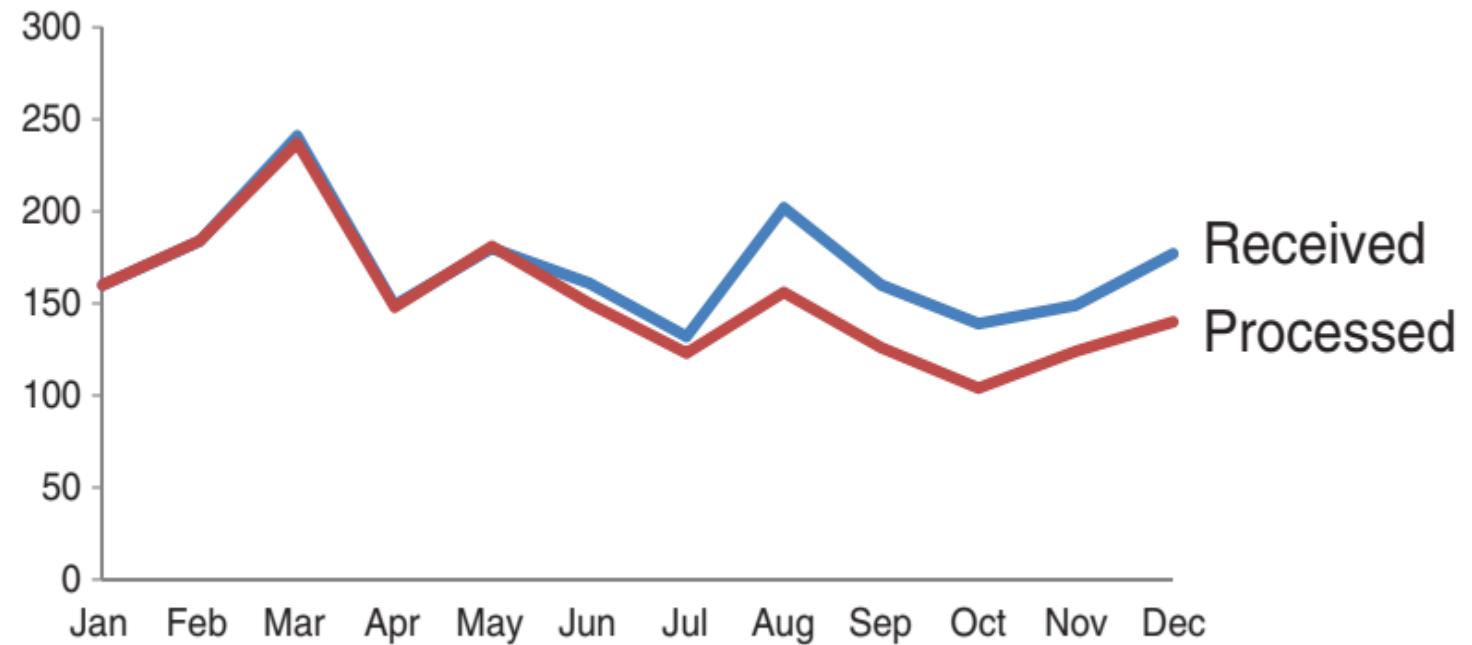


FIGURE 3.22 Label data directly



# Clutter is your enemy

## Decluttering: step-by-step:

1. Remove chart border
2. Remove gridlines
3. Remove data markers
4. Clean up axis labels
5. Label data directly
6. Leverage consistent color

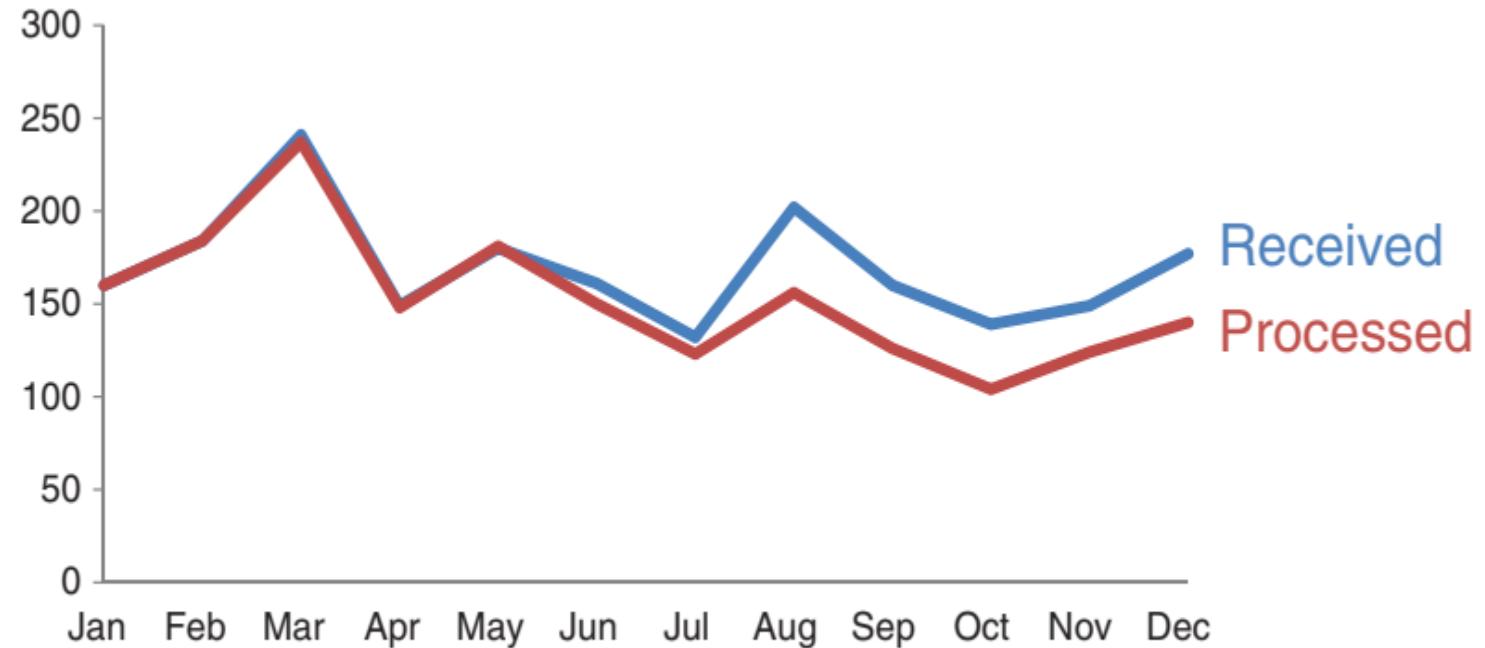


FIGURE 3.23 Leverage consistent color



# Contents

1

Clutter is your enemy

2

Focus your audience's attention

3

Lessons in Storytelling



# DATA VISUALIZATION



# UNIT - II



# Contents

1

**Clutter is your enemy**

2

**Focus your audience's attention**

3

**Lessons in Storytelling**

# How you'll learn to tell stories with data



## STEP 1

### CONTEXT

Understand the Context



## STEP 2

### DISPLAY

Choose an appropriate visual display



## STEP 3

### CLUTTER

Eliminate Clutter



## STEP 4

### ATTENTION

Focus attention where you want it



## STEP 5

### THINKING

Think like a designer



## STEP 6

### TELL A STORY

Tell a story to audience





# Focus your audience's attention

- How people see and how you can use that to your advantage when crafting visuals.
- We will talk briefly about sight and memory to highlight the importance of some specific, powerful tools: preattentive attributes.





# Focus your audience's attention

## You see with your brain:

- Light reflects from a stimulus. This gets captured by our eyes.
- We don't fully see with our eyes; there is some processing that happens there, but mostly it is what happens in our brain that we think of as visual perception.

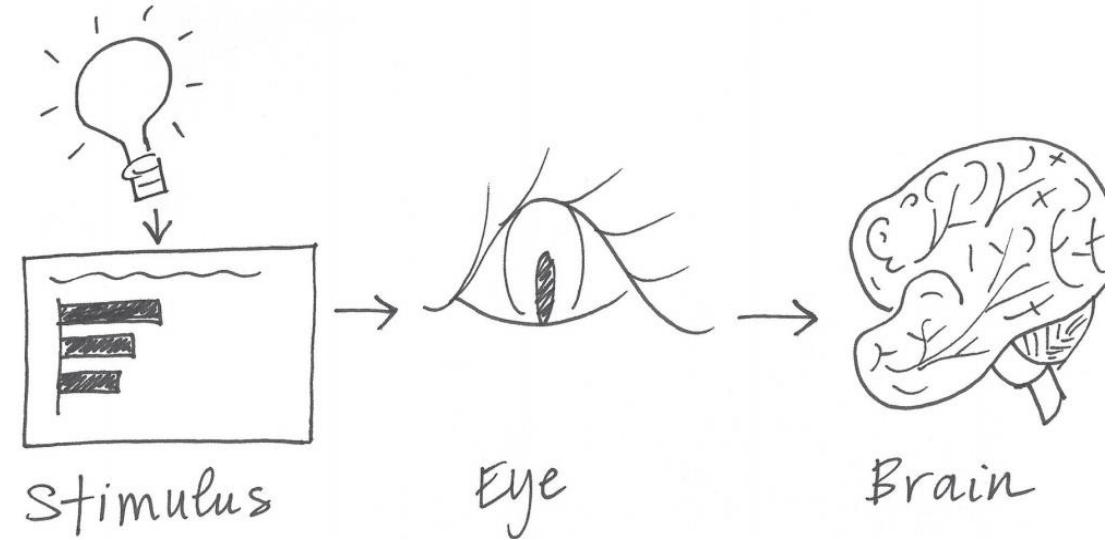


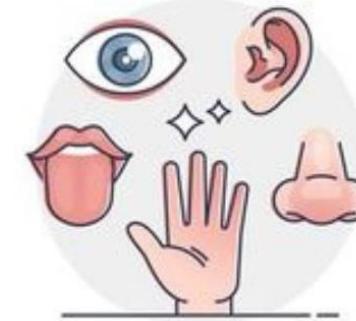
FIGURE 4.1 A simplified picture of how you see



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory



**ICONIC  
MEMORY**



**LONG-TERM  
MEMORY**



**SHORT-TERM  
MEMORY**



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. **Iconic memory**
  2. Short-term memory
  3. Long-term memory

- Iconic memory is super fast.
- It happens without you consciously realizing it.
- For example, when you see a car passing by on the highway, and for a moment you can picture the car after it is gone.
- Information stays in your iconic memory for a fraction of a second before it gets forwarded on to your short-term memory.



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory





# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory

- Short-term memory has limitations.
- Specifically, people can keep about four chunks of visual information in their short-term memory at a given time.



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory
- Remembering a phone number
- Remembering details from a book you read a few days ago
- Following directions
- Listening to a lecture



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory

- When something leaves short-term memory, it either goes into oblivion and is likely lost forever or is passed into long-term memory.
- Long-term memory is built up over a lifetime and is vitally important for pattern recognition and general cognitive processing.



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory

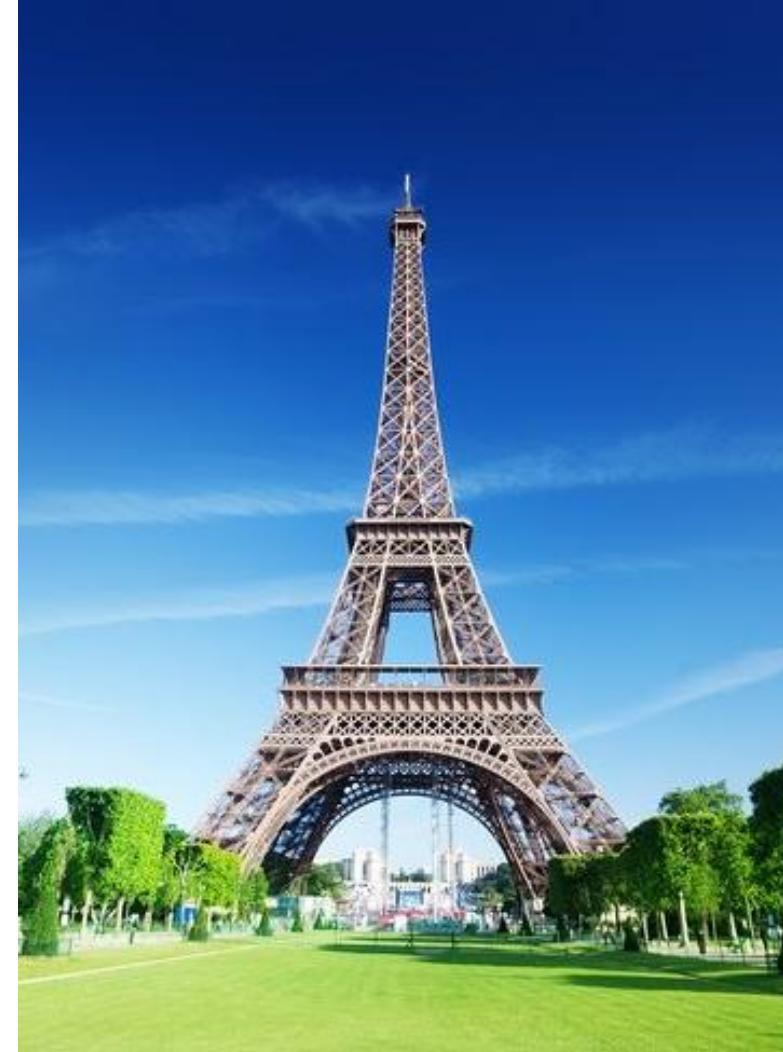
- Remembering important days, such as birthdays or holidays
- Knowing how to ride a bicycle
- Remembering the words to a song
- Knowing how to type
- Recollecting work skills learned in your first job



# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory





# Focus your audience's attention

## A brief lesson on memory:

- There are three types of memory that are important to understand as we design visual communications:
  1. Iconic memory
  2. Short-term memory
  3. Long-term memory











# Focus your audience's attention

Preattentive attributes signal where to look:

- Figure 4.2 shows a block of numbers. Taking note of how you process the information and how long it takes, quickly count the number of 3s that appear in the sequence.

756395068473  
658663037576  
860372658602  
846589107830

FIGURE 4.2 Count the 3s example



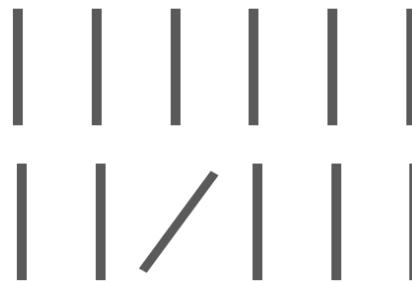
# Focus your audience's attention

Preattentive attributes signal where to look:

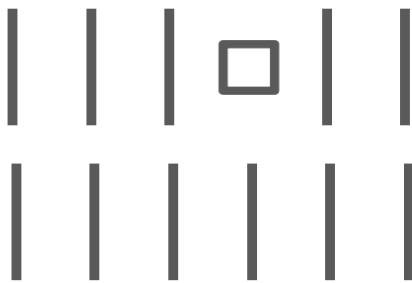
- Use preattentive attributes strategically, they can help us enable our audience to see what we want them to see before they even know they're seeing it!

756395068473  
658663037576  
860372658602  
846589107830

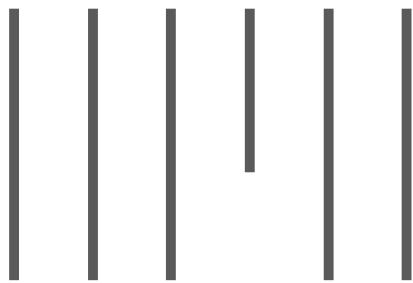
FIGURE 4.3 Count the 3s example with preattentive attributes



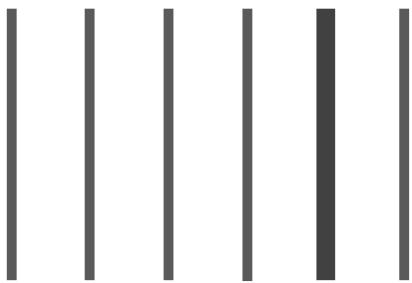
Orientation



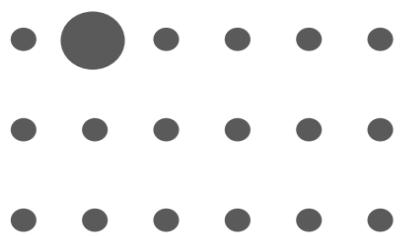
Shape



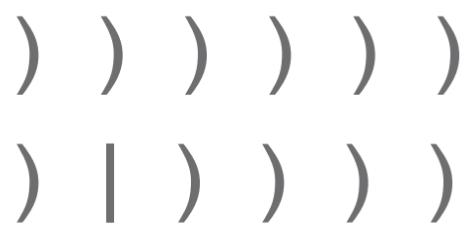
Line length



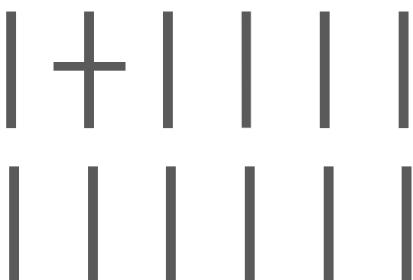
Line width



Size



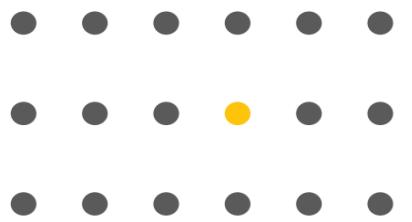
Curvature



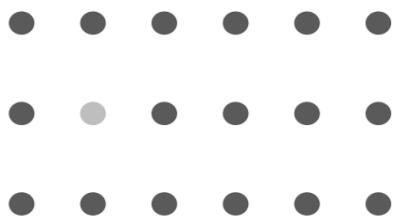
Added marks



Enclosure



Hue



Intensity



Spatial position



Motion

**FIGURE 4.4** Preattentive attributes



# Focus your audience's attention

## Preattentive attributes in text:

- The subsequent blocks of text employ a single preattentive attribute each.
- Note how, within each, the preattentive attribute grabs your attention, and how some attributes draw your eyes with greater or weaker force than others.

## No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# Bold

**What are we doing well?** Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# Color

What are we doing well? Great Products. These products are clearly the best in their class.

Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# Italics

What are we doing well? Great Products. These products are clearly the best in their class.

*Replacement parts are shipped when needed.* You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets

**without me having to ask.** Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a

# Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class.

Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly.

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a

# Outline (enclosure)

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# Underline (added marks)

What are we doing well? Great Products. These products are clearly the best in their class.

Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.

You have a great company – keep up the good work!

# What are we doing well?

Themes & example comments

- **Great products:** "These products are clearly the best in class."
- **Replacement parts are shipped when needed:**  
"You sent me gaskets without me having to ask, and I really needed them, too!"
- **Problems are resolved promptly:** "Bev in the billing office was quick to resolve a billing issue I had."
- **General customer service exceeds expectations:**  
"The account manager even called after normal business hours.  
*You have a great company - keep up the good work!*"

**FIGURE 4.6** Preattentive attributes can help create a visual hierarchy of information



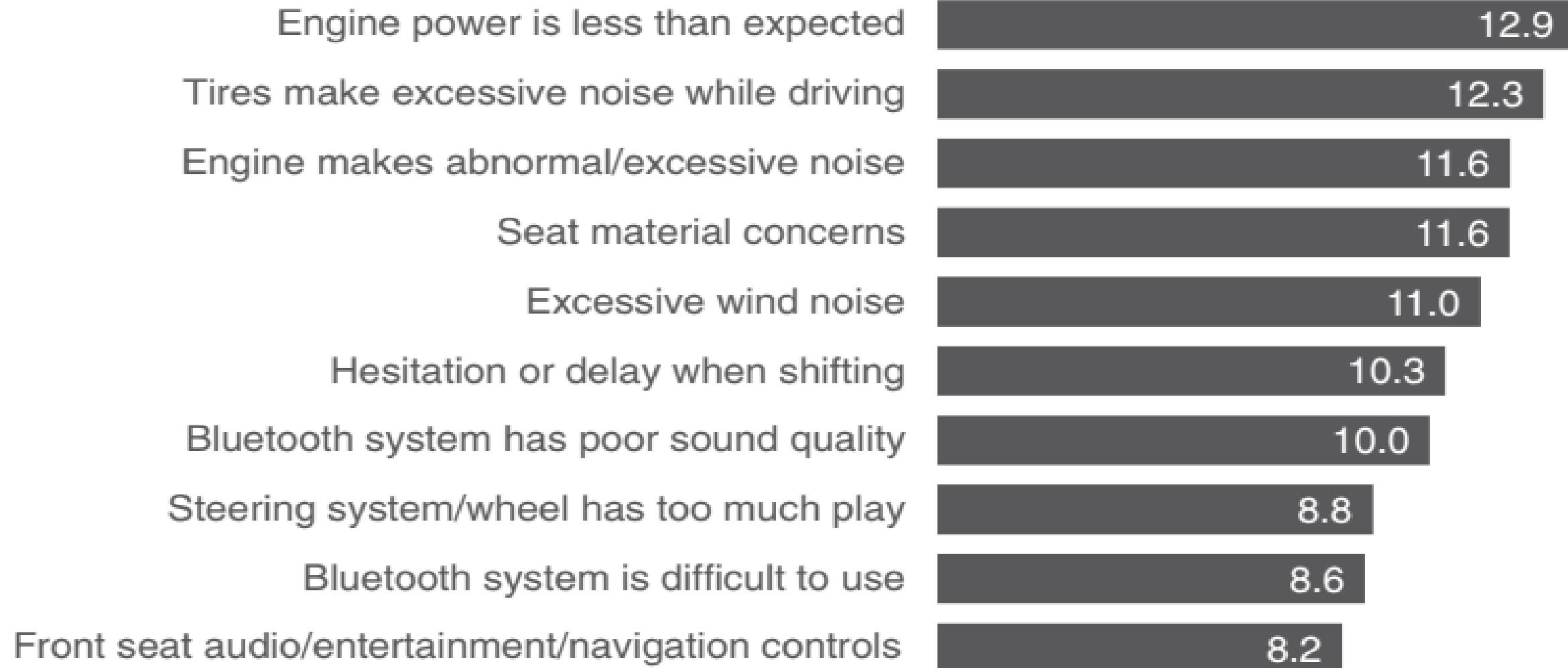
# Focus your audience's attention

## Preattentive attributes in graphs:

- Imagine you work for a car manufacturer.
- You are interested in understanding and sharing insight about the top design concerns from customers for a particular vehicle make and model.
- Measured as the number of concerns per 1,000 concerns from customers for a particular vehicle.
- Your initial visual might look something like Figure 4.7

## Top 10 design concerns

concerns per 1,000

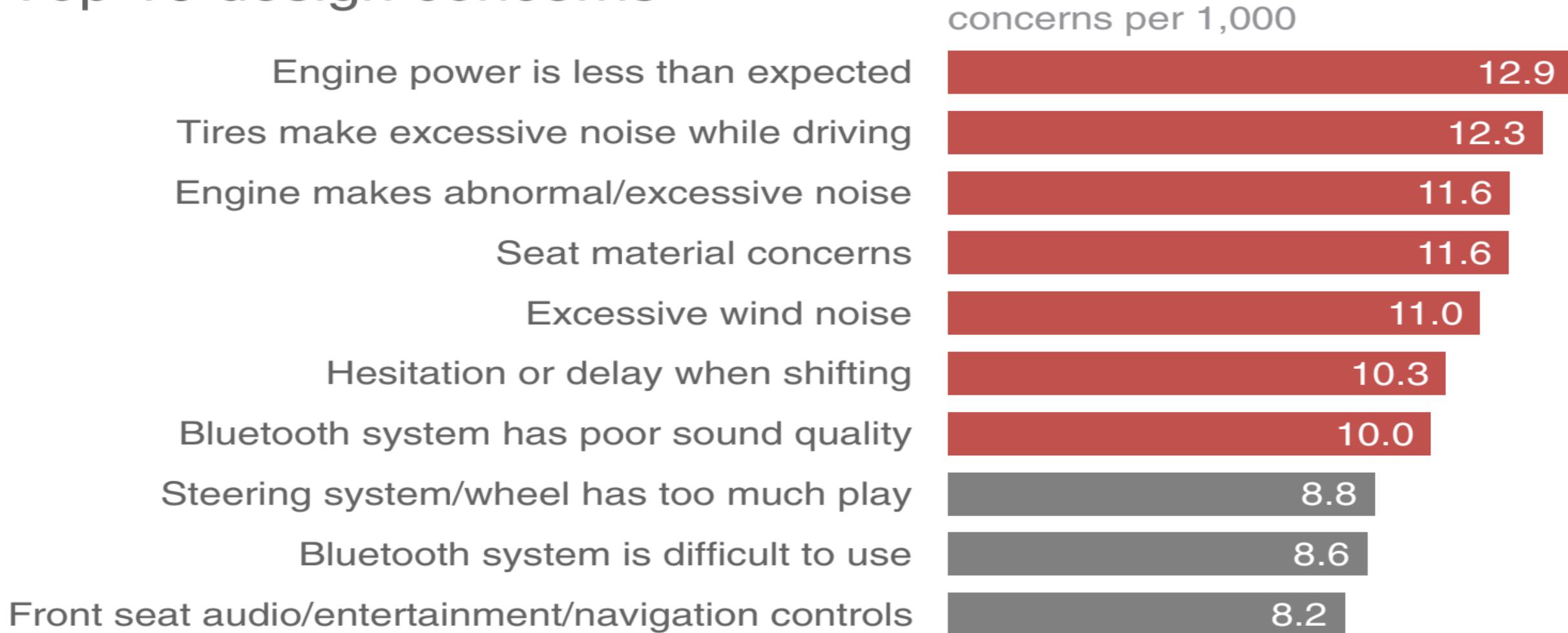


**FIGURE 4.7** Original graph, no preattentive attributes

**7 of the top 10 design concerns have 10 or more concerns per 1,000.**

Discussion: is this an acceptable default rate?

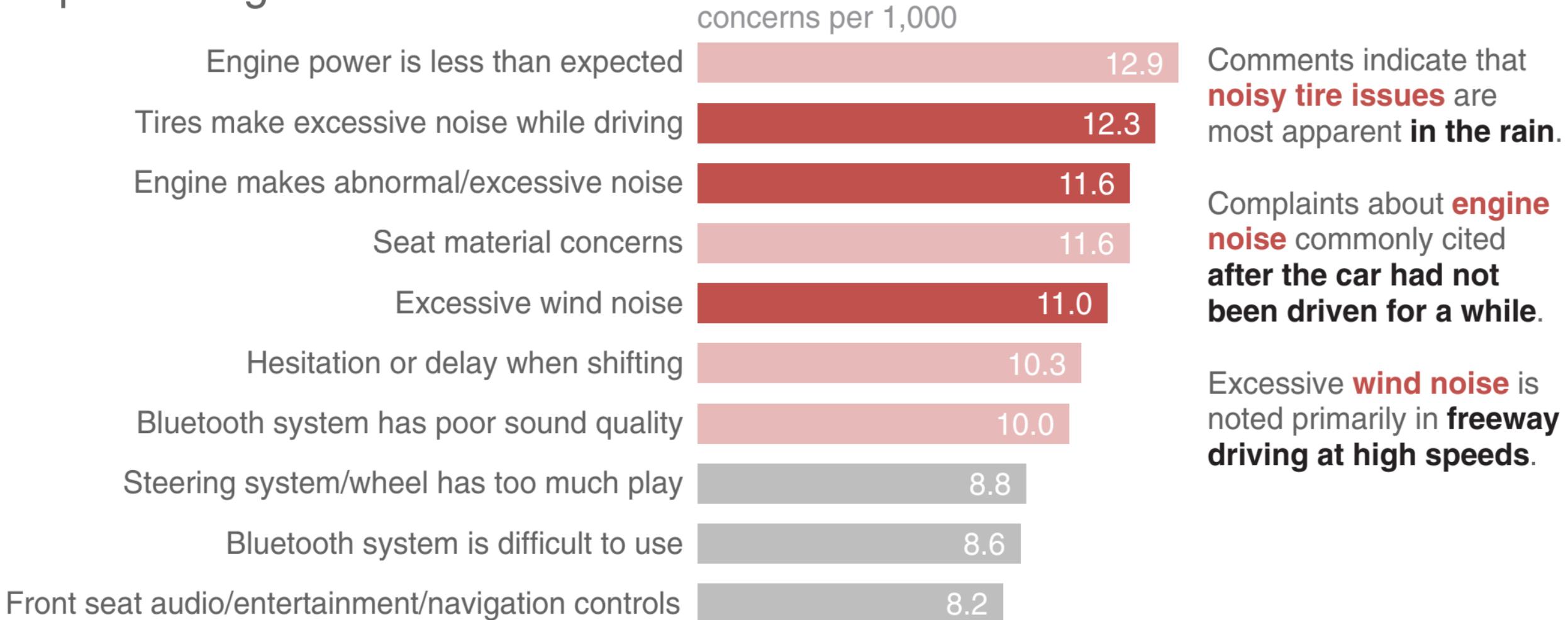
## Top 10 design concerns



**FIGURE 4.8** Leverage color to draw attention

## Of the top design concerns, three are noise-related.

### Top 10 design concerns



**FIGURE 4.9** Create a visual hierarchy of information



# Focus your audience's attention

## Preattentive attributes in graphs:

- There are a few preattentive attributes that are so important from a strategic standpoint when it comes to focusing your audience's attention that they warrant their own specific discussions:

**size, color, and position on page.**



# Focus your audience's attention

- **Size**
- Color
- Position

- If you're showing multiple things that are of roughly equal importance, size them similarly.
- Alternatively, if there is one important thing, leverage size to indicate that: make it BIG!



# Focus your audience's attention

- Size
- **Color**
- Position

- When used sparingly, color is one of the most powerful tools you have for drawing your audience's attention.
- Resist the need to use color for the sake of being colorful; instead, leverage color selectively as a strategic tool to highlight the important parts of your visual.



# Focus your audience's attention

- Size
- **Color**
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- Brand colors.



# Focus your audience's attention

- Size
- Color
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- Brand colors.



# Focus your audience's attention

- Size
- Color
- Position

- A table that showed market rank for a handful of pharmaceutical drugs across a few different countries.
- Each rank (1, 2, 3, and so on) was assigned its own color along a rainbow spectrum: 1 = red, 2 = orange 3 = yellow, 4 = light green, 5 = green, 6 = teal, 7 = blue, 8 = dark blue, 9 = light purple, 10+ = purple.



# Focus your audience's attention

- Size
- Color
- Position

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5



# Focus your audience's attention

- Size
- Color
- Position

Top 5 drugs: country-level sales rank

RANK	1	2	3	4	5+
COUNTRY   DRUG	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5



# Focus your audience's attention

- Size
- Color
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- Brand colors.



# Focus your audience's attention

- Size
- **Color**
- Position

- There are some cases where use of color must be consistent.
- Your audience will typically take time to familiarize themselves with what colors mean once and then will assume the same details apply throughout the rest of the communication.



# Focus your audience's attention

- Size
- **Color**
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- Brand colors.



# Focus your audience's attention

- Size
  - Color
  - Position
- 
- This most frequently manifests itself as difficulty in distinguishing between shades of red and shades of green.
  - When designing a visual and selecting colors to highlight both positive and negative aspects, frequently use blue to signal positive and orange for negative.



# Focus your audience's attention

- Size
- **Color**
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- Brand colors.



# Focus your audience's attention

- Size
- Color
- Position

- I had created these visuals in my typical color palette: shades of grey with a medium blue used sparingly to draw attention.



# Focus your audience's attention

- Size
- **Color**
- Position

When it comes to the use of color, there are several specific lessons to know:

- Use it sparingly
- Use it consistently
- Design with the colorblind in mind
- Be thoughtful of the tone color conveys
- **Brand colors.**



# Focus your audience's attention

- Size
- **Color**
- Position

- I was once working with a client whose brand color was a light shade of green. I originally wanted to leverage this green as the standout color, but it simply wasn't attention grabbing enough.
- There wasn't sufficient contrast, so the visuals I created had a washed-out feel.



# Focus your audience's attention

- Size
- Color
- Position

Leverage **brand color**

Category 1 7

Category 2 5

Category 3 4

Category 4 4

Category 5 3

*ClientLogo*

Draw attention with **black**

Category 1 7

Category 2 5

Category 3 4

Category 4 4

Category 5 3

*ClientLogo*

Use **complementary color**

Category 1 7

Category 2 5

Category 3 4

Category 4 4

Category 5 3

*ClientLogo*

FIGURE 4.16 Color options with brand color



# Focus your audience's attention

- Size
- Color
- Position

- Most members of your audience will start at the top left of your visual or slide and scan with their eyes in zigzag motions across the screen or page.

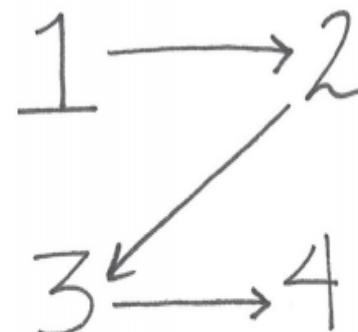


FIGURE 4.17 The zigzag "z" of taking in information on a screen or page



# Contents

1

**Clutter is your enemy**

2

**Focus your audience's attention**

3

**Lessons in Storytelling**



# DATA VISUALIZATION



# UNIT - II



# Contents

1

Clutter is your enemy

2

Focus your audience's attention

3

Lessons in Storytelling

# How you'll learn to tell stories with data



## STEP 1

### CONTEXT

Understand the Context



## STEP 2

### DISPLAY

Choose an appropriate visual display



## STEP 3

### CLUTTER

Eliminate Clutter



## STEP 4

### ATTENTION

Focus attention where you want it



## STEP 5

### THINKING

Think like a designer



## STEP 6

### TELL A STORY

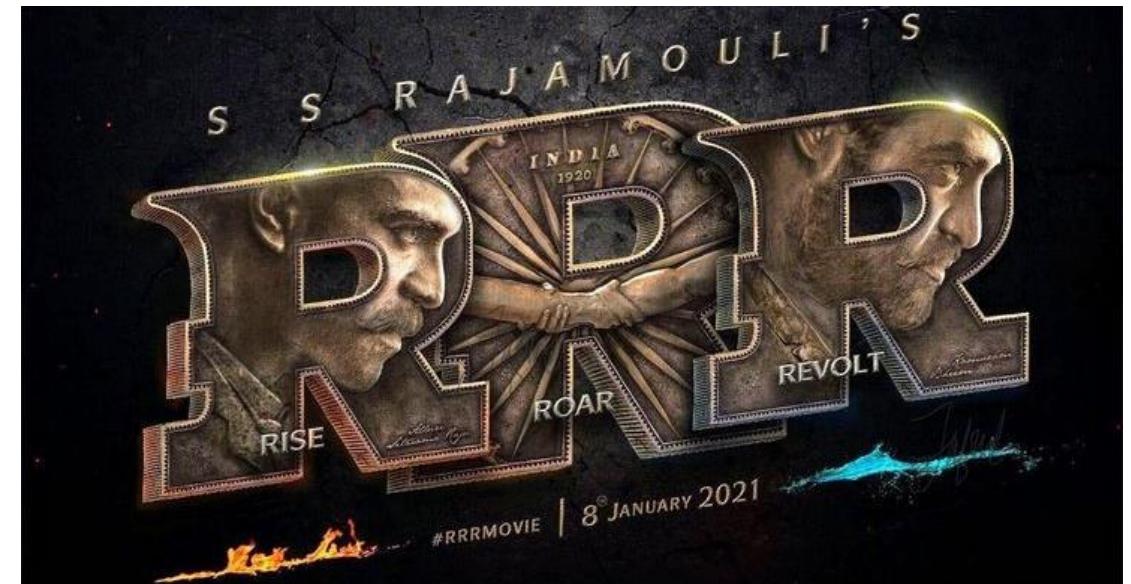
Tell a story to audience





# Lessons in Storytelling

- The lesson on storytelling often begins with a thought exercise.
- Close their eyes and recall the story of “RRR”





# Lessons in Storytelling

- The lesson on storytelling often begins with a thought exercise.
- Close their eyes and recall the story of “EVARU”





# Lessons in Storytelling

- The lesson on storytelling often begins with a thought exercise.
- Close their eyes and recall the story of “THE MUMMY”





# Lessons in Storytelling

## The magic of story:

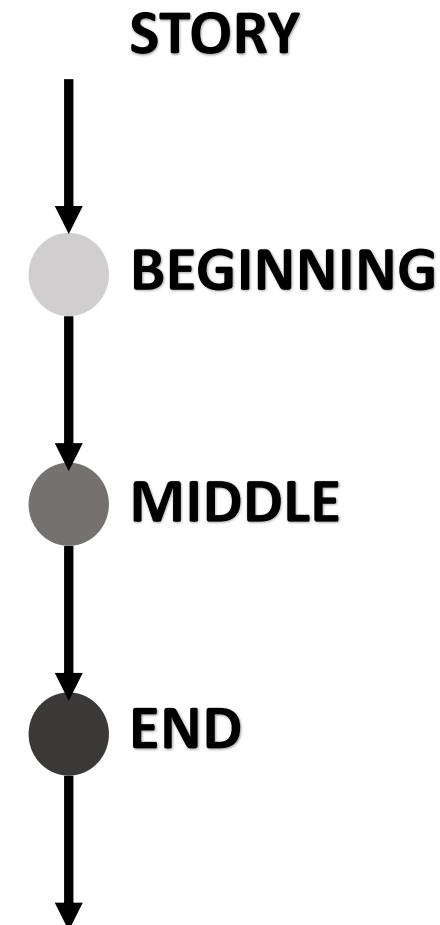
- When you see a great play, watch a captivating movie, or read a fantastic book, you've experienced the magic of story.
- A good story grabs your attention and takes you on a journey, evoking an emotional response.
- In the middle of it, you find yourself not wanting to turn away or put it down.
- After finishing it—a day, a week, or even a month later—you could easily describe it to a friend



# Lessons in Storytelling

## Storytelling in plays or Cinema:

- Aristotle introduced a basic but profound idea: that story has a clear beginning, middle, and end.
- He proposed a three-act structure for plays.





# Lessons in Storytelling

## Storytelling in plays or Cinema:

The first act sets up the story (The Beginning).

- The first thing to do is introduce the plot, building the context for your audience.
- In this section, we set up the essential elements of story—the setting, main character, unresolved situation, and desired outcome getting everyone on common ground so the story can proceed.





# Lessons in Storytelling

## Storytelling in plays or Cinema:

We should involve our audience

- Why should audience pay attention?
- What is in it for audience?





# Lessons in Storytelling

## Storytelling in plays or Cinema:

- It introduces the main character, or hero, their relationships, and the world in which they live.
  1. **The setting:** When and where does the story take place?
  2. **The main character:** Who is driving the action? (This should be framed in terms of your audience!)
  3. **The imbalance:** Why is it necessary, what has changed?
  4. **The balance:** What do you want to see happen?
  5. **The solution:** How will you bring about the changes?





# Lessons in Storytelling

## Storytelling in plays or Cinema:

The second act makes up the bulk of the story (The middle).

- Once you've set the stage, so to speak, the bulk of your communication further develops "what could be," with the goal of convincing your audience of the need for action. You retain your audience's attention through this part of the story by addressing how they can solve the problem you introduced.





# Lessons in Storytelling

## Storytelling in plays or Cinema:

- The main character lacks the skills to deal with the problem he faces, and, as a result, finds himself encountering increasingly worsening situations.
- He may have to learn new skills or reach a higher sense of awareness of who he is and what he is capable of to deal with his situation.





# Lessons in Storytelling

## Storytelling in plays or Cinema:

The third act resolves the story and its subplots (The End).

- Finally, the story must have an end. End with a call to action: make it totally clear to your audience what you want them.
- It includes a climax, where the tensions of the story reach the highest point of intensity.





# Lessons in Storytelling

## The narrative structure:

- To be successful, a narrative must be central to the communication.
- These are words—written, spoken, or a combination of the two—that tell the story in an order that makes sense and convinces the audience why it's important or interesting and attention to it should be paid.

## Considerations:

- Let's discuss some specific considerations when it comes to both the order of the story and the spoken and written narrative.



# Lessons in Storytelling

## Narrative flow: the order of your story:

- Your story must have an order to it.
- A collection of numbers and words on a given topic without structure to organize them and give them meaning is useless.

- One way to order the story—the one that typically comes most naturally—is **chronologically**.
- Another strategy is to **lead with the ending**.



# Lessons in Storytelling

## The spoken and written narrative:

- With a live presentation, you have the benefit of words on the screen or page being reinforced by the words you are saying.
- In this manner, your audience could both read and hear what they need to know, strengthening the information.

- Another challenge is that your audience can act unpredictably.
- They can ask questions that are off topic, jump to a point later in the presentation, or do other things to push you off track.



# Lessons in Storytelling

## The spoken and written narrative:

- In a written report you don't have the benefit of the voiceover to make the sections or slides relevant rather, they must do this on their own.
- Think about what words need to be present.

- Getting feedback from someone not as familiar with the topic can be especially useful in this situation.
- Doing so will help you uncover issues with clarity and flow, or questions your audience may have, so you can address those proactively.



# Lessons in Storytelling

## The power of repetition:

- Thinking back to “RRR”, one of the reasons we remember the story is due to repetition.
- The more the information is repeated or used, the more likely it is to eventually end up in long-term memory, or to be retained.

- Let’s explore a concept called **Bing, Bang, Bongo.**
- The idea is that you should first tell your audience what you’re going to tell them (“Bing,” the introduction paragraph in your essay).
- Then you tell it to them (“Bang,” the actual essay content).
- Then you summarize what you just told them (“Bongo,” the conclusion).



# Lessons in Storytelling

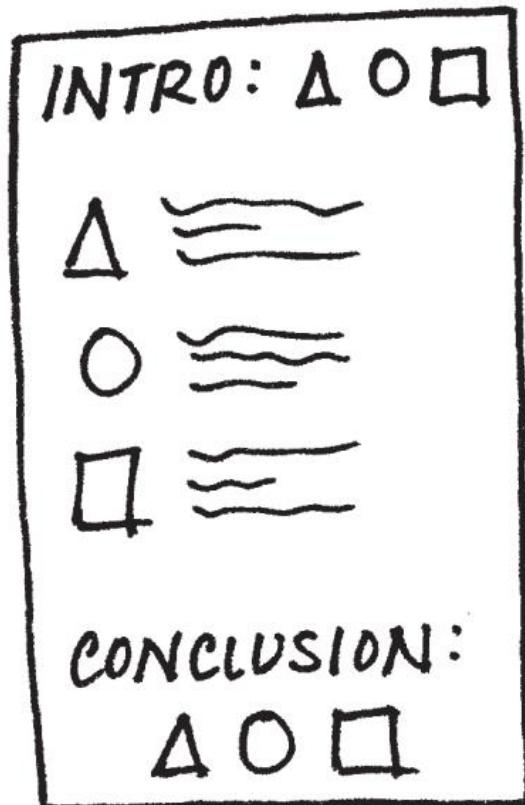


FIGURE 7.1 Bing, bang, bongo

- Let's explore a concept called **Bing, Bang, Bongo**.
- The idea is that you should first tell your audience what you're going to tell them ("Bing," the introduction paragraph in your essay).
- Then you tell it to them ("Bang," the actual essay content).
- Then you summarize what you just told them ("Bongo," the conclusion).



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

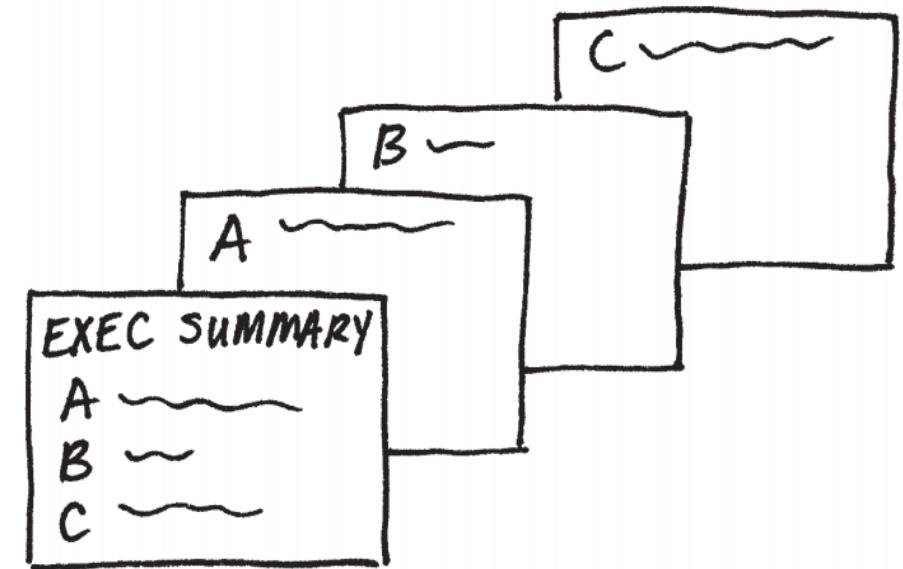


FIGURE 7.2 Horizontal logic



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

- One strategy is to have an executive summary slide up front, with each bullet corresponding to a subsequent slide title in the same order (Figure 7.2).



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. **Vertical logic**
3. Reverse storyboarding
4. Fresh perspective

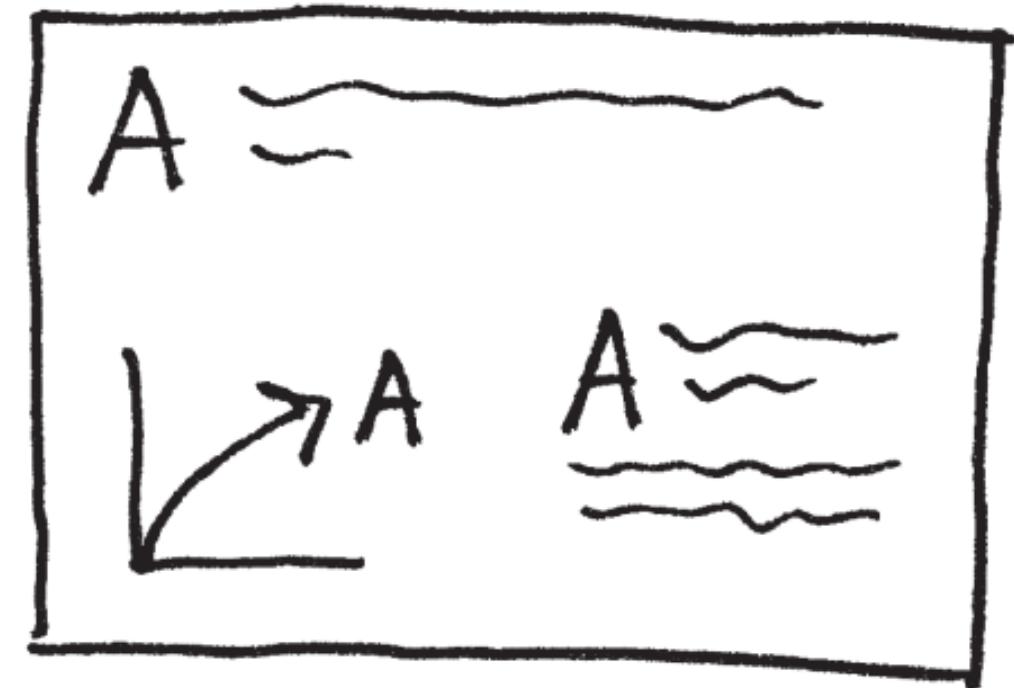


FIGURE 7.3 Vertical logic



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

- Vertical logic means that all information on a given slide is self-reinforcing. The content reinforces the title and vice versa. The words reinforce the visual and vice versa (Figure 7.3).
- There isn't any extraneous or unrelated information.



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

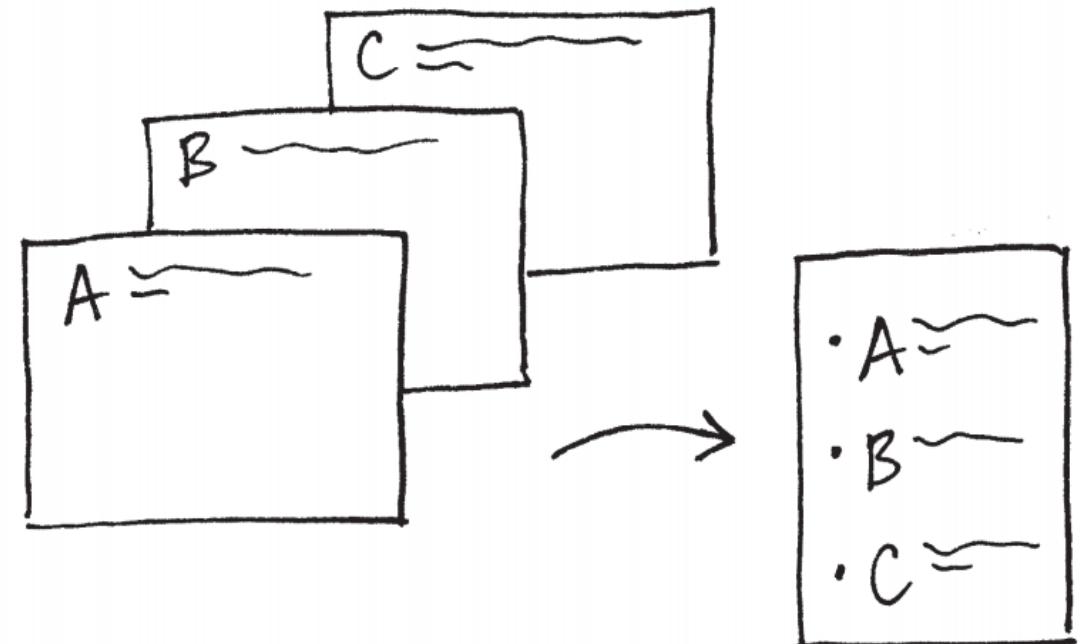


FIGURE 7.4 Reverse storyboarding



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

- You take the final communication, flip through it, and write down the main point from each page.
- Resulting list should look like the storyboard or outline for the story you want to tell (Figure 7.4).



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

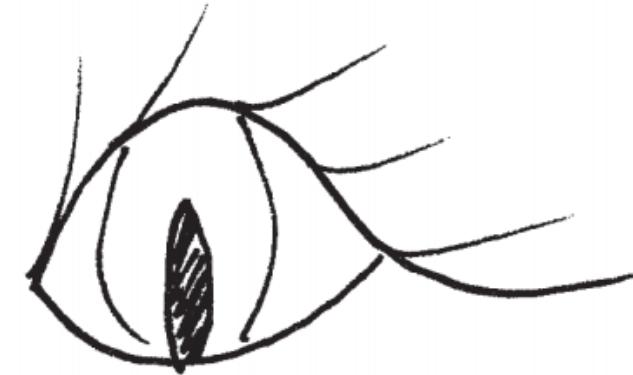


FIGURE 7.5 A fresh perspective



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

- Once you've crafted your communication, give it to a friend or colleague. It can be someone without any context.



# Lessons in Storytelling

Tactics to help ensure that your story is clear:

Four Tactics:

1. Horizontal logic
2. Vertical logic
3. Reverse storyboarding
4. Fresh perspective

- Ask them to tell you what they pay attention to, what they think is important, and where they have questions.
- This will help you understand whether the communication you've crafted is telling the story you mean to tell .



# Contents

1

**Clutter is your enemy**

2

**Focus your audience's attention**

3

**Lessons in Storytelling**



# DATA VISUALIZATION



# **UNIT - III (PART - I)**



# Communicating data

1

A step in the process

2

A model of communication

3

Communication problems

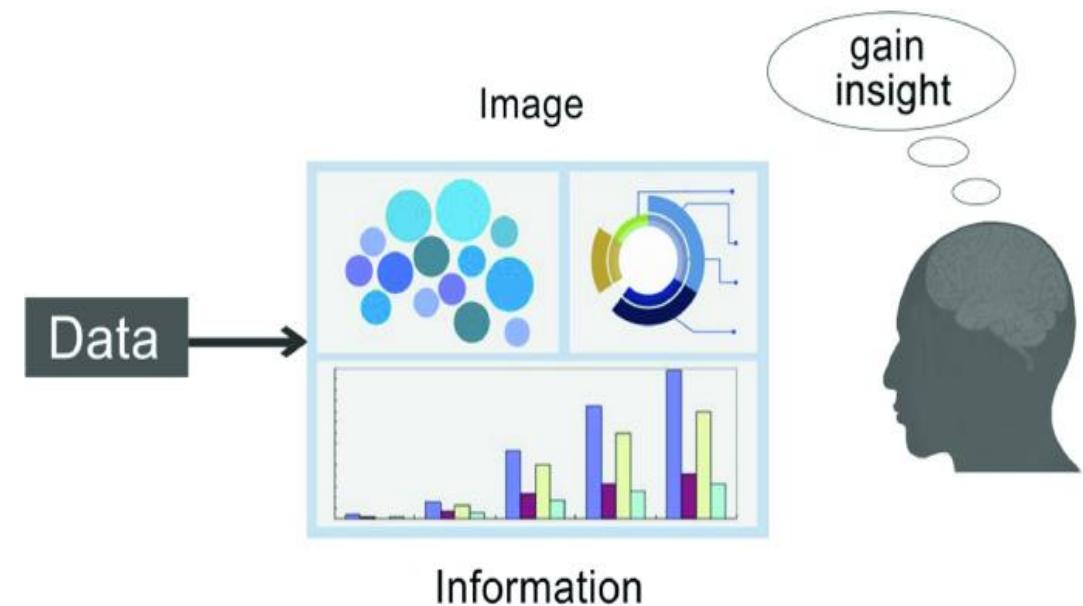
4

Principles of communicating data



# A step in the process

- When data is communicated well, it's easy to appreciate both the data itself and the delivery of that data at the same time.
- Those two elements can be fashioned together into an overall experience that makes you feel that you understand the world better

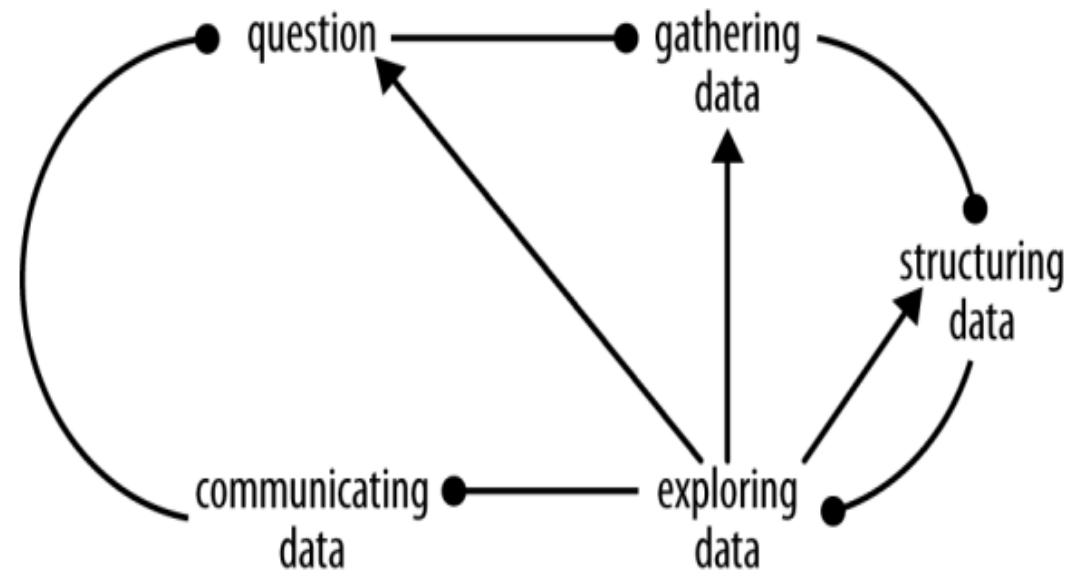


©



# A step in the process

- How is “communicating data” distinct from the other steps in the overall process that begins with a question and ends with a shared insight?
- Figure 1-1 presents the overall data discovery process and shows where communicating data fits in that process.



*Figure 1-1. The data discovery process*

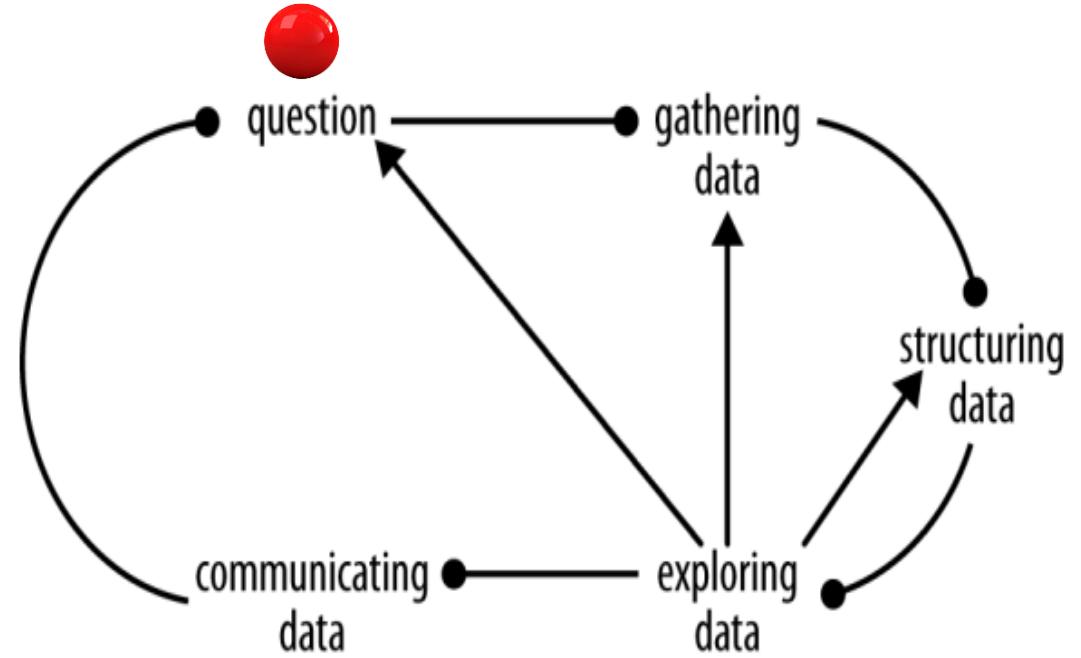


# A step in the process

- The highly iterative process often begins with a **question**, which can be specific or general.

**Ex:**

- which combination of products occurs the most often? (Specific)
- what can we learn about historical sales of our products? (General)

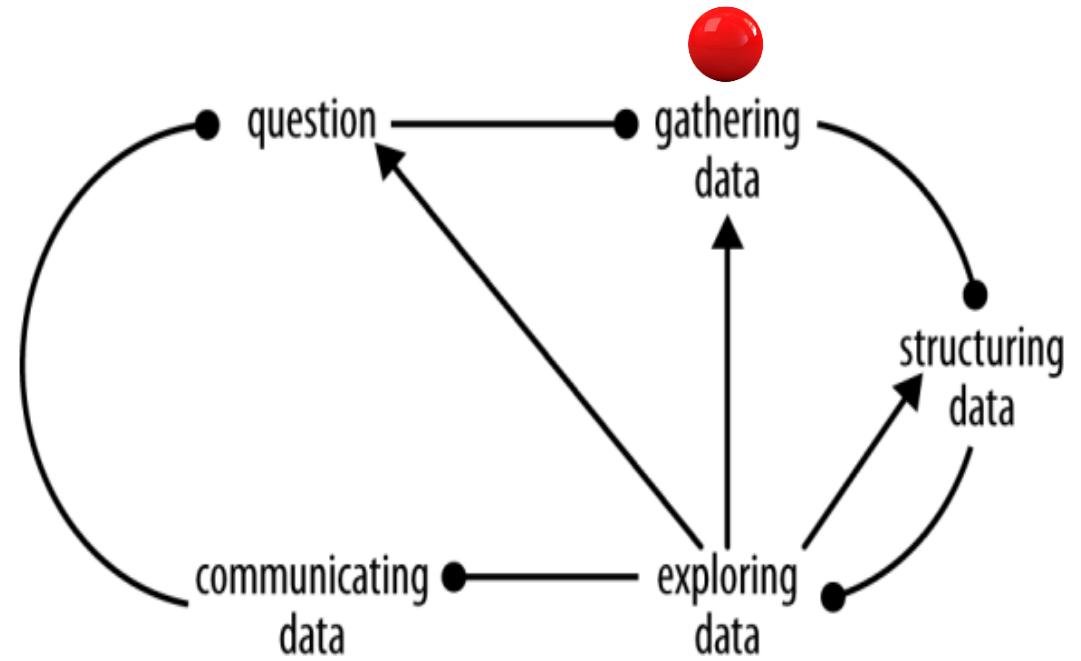


*Figure 1-1. The data discovery process*



# A step in the process

- The next step is **gathering data** if it's available (e.g., historical sales).

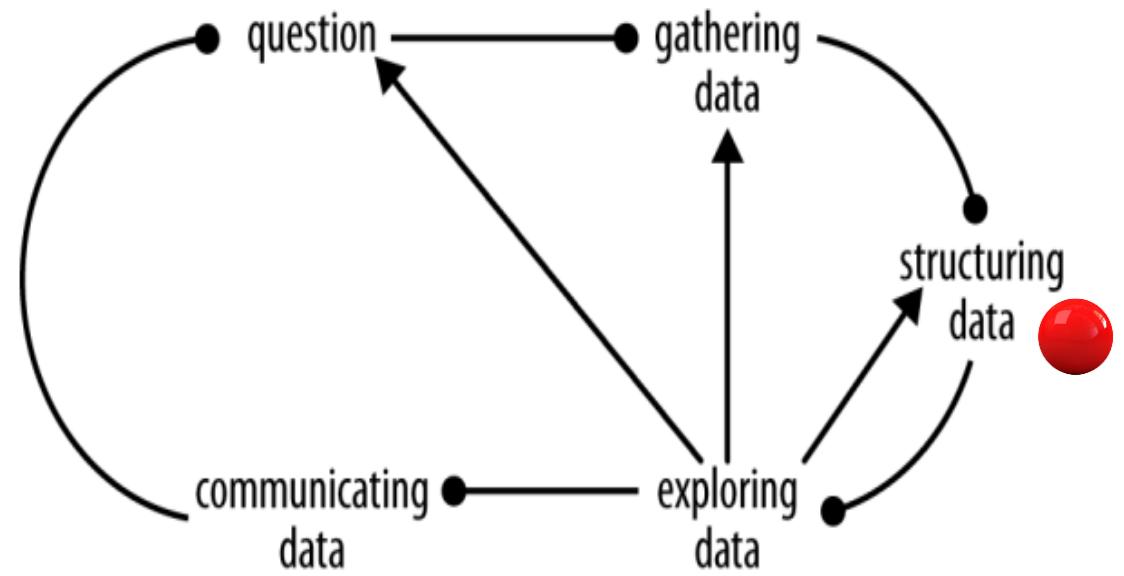


*Figure 1-1. The data discovery process*



# A step in the process

- Then comes the often-difficult process of **structuring data**, also called “data munging” or “data wrangling.”
- In this step, data is formatted, shaped, merged, converted, and otherwise manipulated into a form that is amenable to the next step

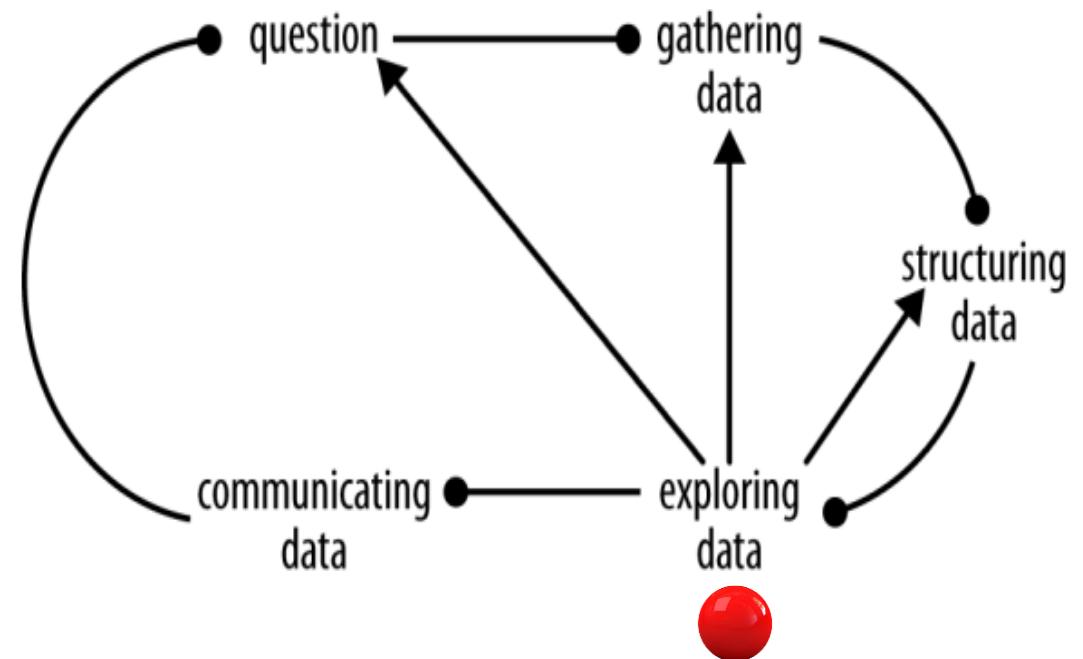


*Figure 1-1. The data discovery process*



# A step in the process

- In this **exploring data** step, the data is viewed and analyzed from several angles until one or more insights are gathered.

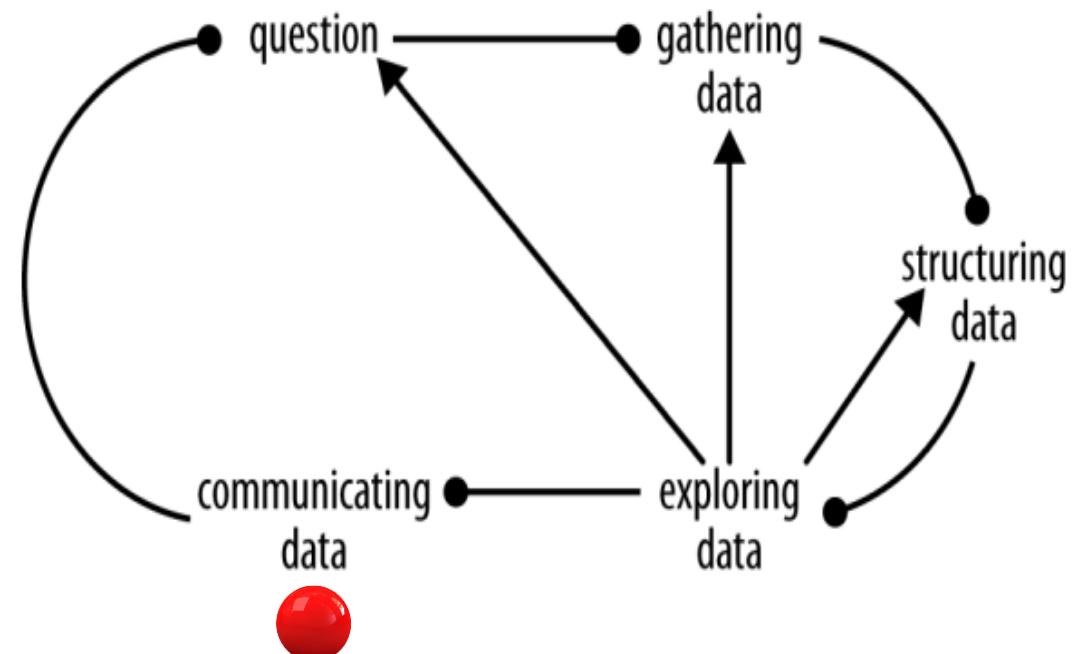


*Figure 1-1. The data discovery process*



# A step in the process

- In **communicating data** step at which quantitative statements are shared with others.



*Figure 1-1. The data discovery process*



# Communicating data

1

A step in the process

2

A model of communication

3

Communication problems

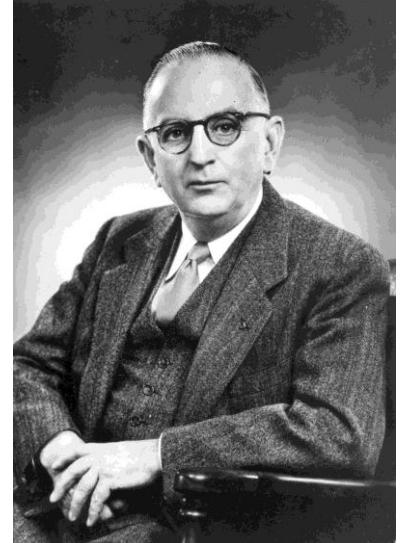
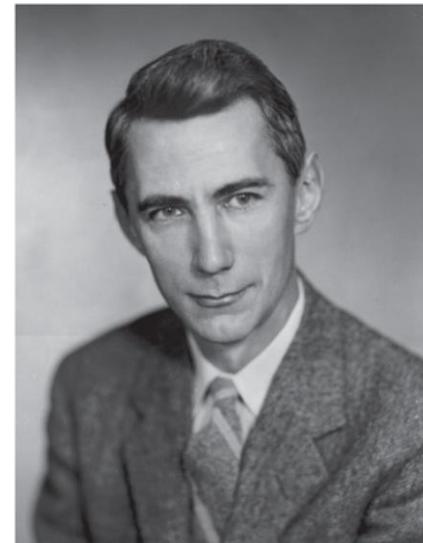
4

Principles of communicating data



# A model of communication

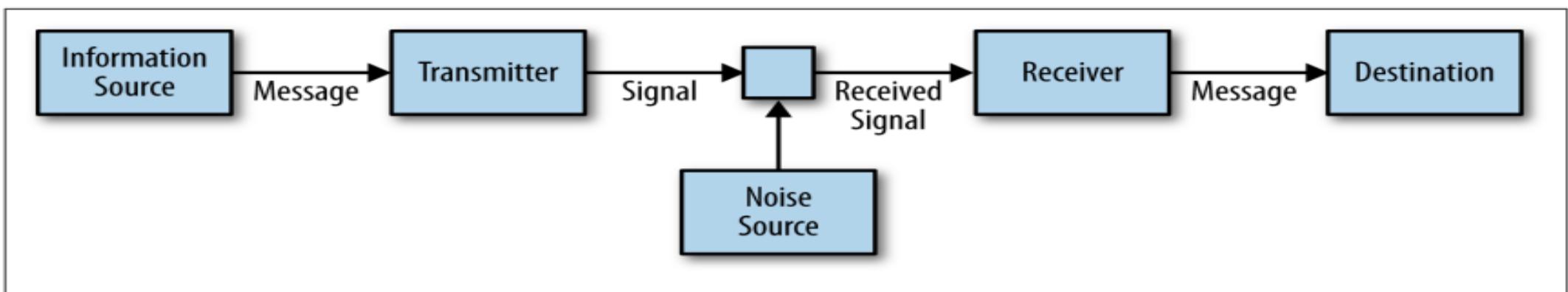
- The year was 1949, and two employees at Bell Laboratories, Claude Elwood Shannon and his coauthor Warren Weaver—published a seminal article in the University of Illinois Press called The Mathematical Theory of Communication.





# A model of communication

- They introduced a model of communication systems in which an “information source” selects a message and then a “transmitter changes this message into the signal which is sent over the communication channel from the transmitter to the receiver” (see Figure 1~2).



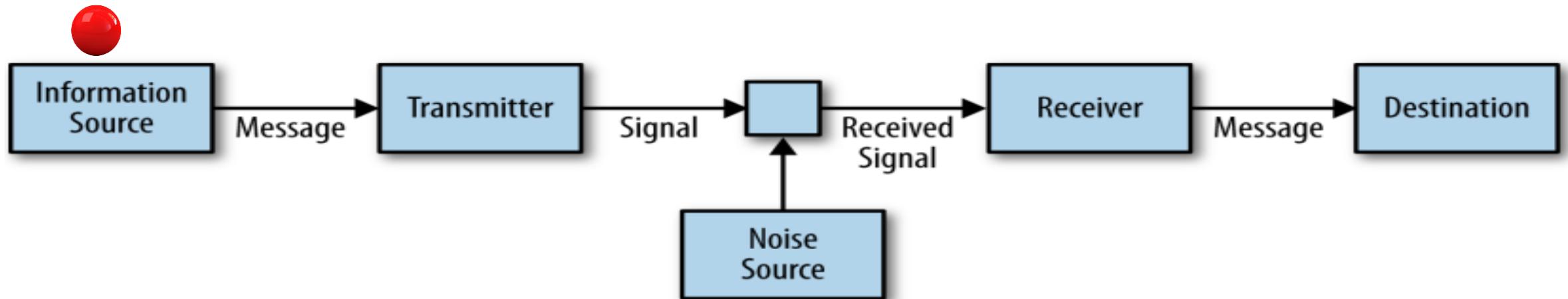
*Figure 1-2. A model of communication systems*



# A model of communication

## Ex: Oral Speech

- The information source is the brain of a certain person

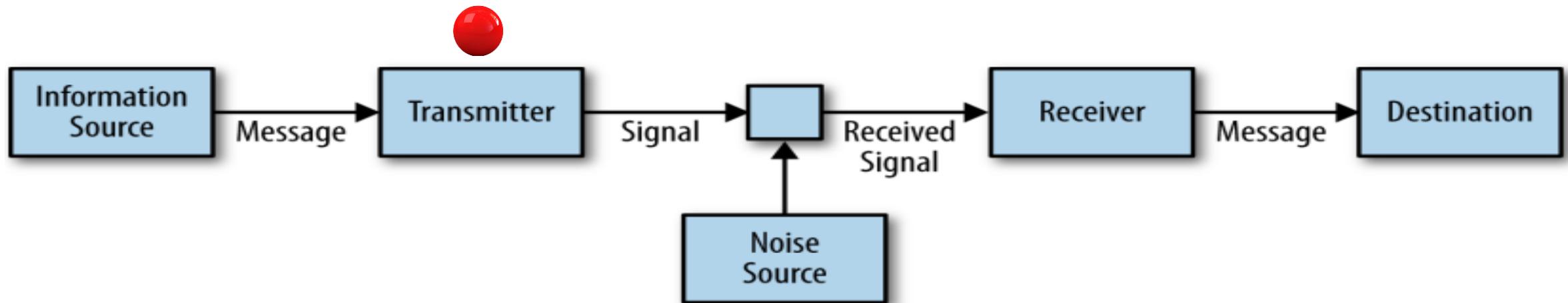




# A model of communication

## Ex: Oral Speech

- The transmitter is this person's vocal system

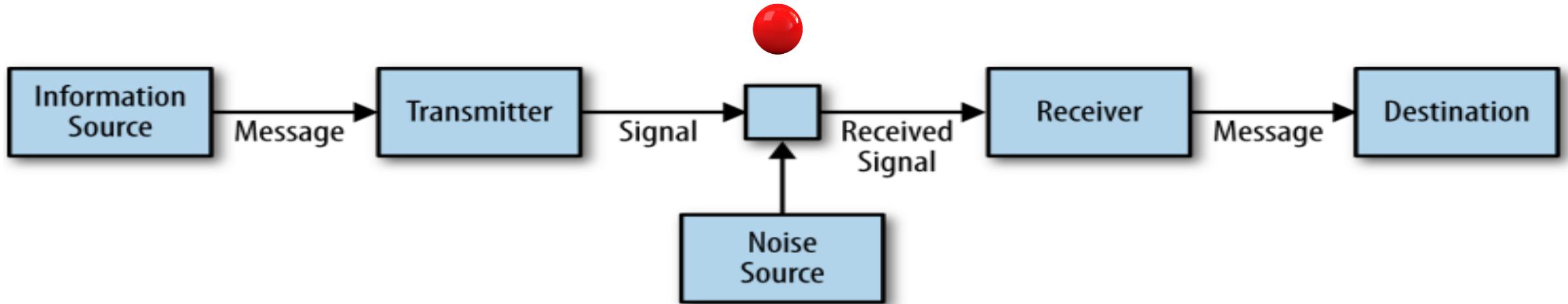




# A model of communication

## Ex: Oral Speech

- The channel is the sound waves that travel as particles in the air collide.

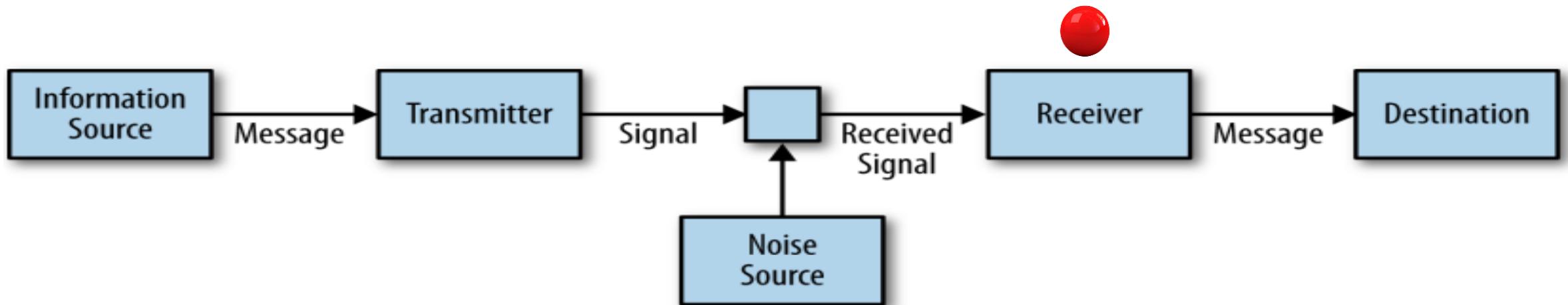




# A model of communication

## Ex: Oral Speech

- The receiver is the **auditory system** of a second person.

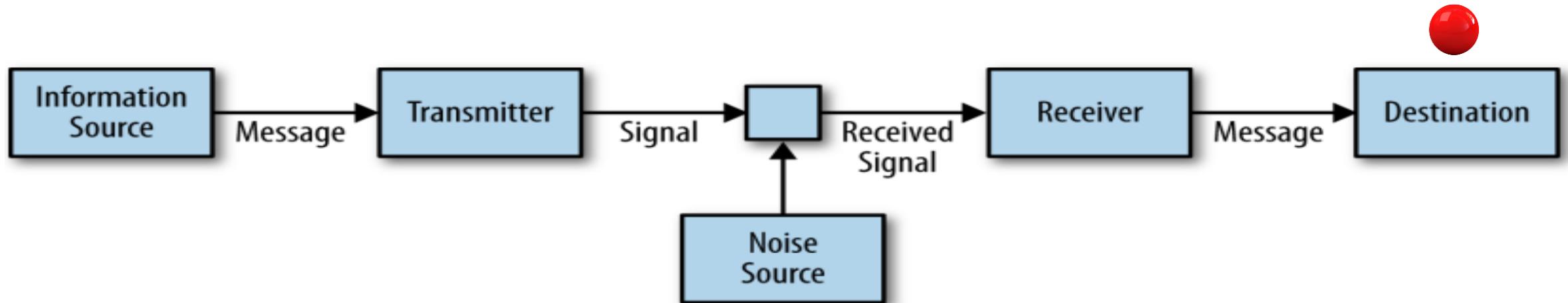




# A model of communication

## Ex: Oral Speech

- The destination is this second person's brain

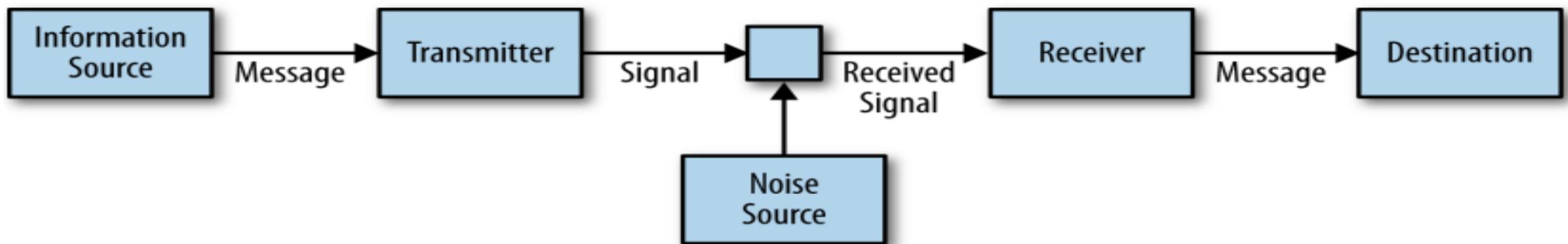




# A model of communication

## Ex: Oral Speech

- The noise source includes other sounds present at the time the first person speaks.





# Communicating data

1

A step in the process

2

A model of communication

3

Communication problems

4

Principles of communicating data



# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

**Three  
Problems**



# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

- How accurately can the symbols of communication be transmitted?





# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

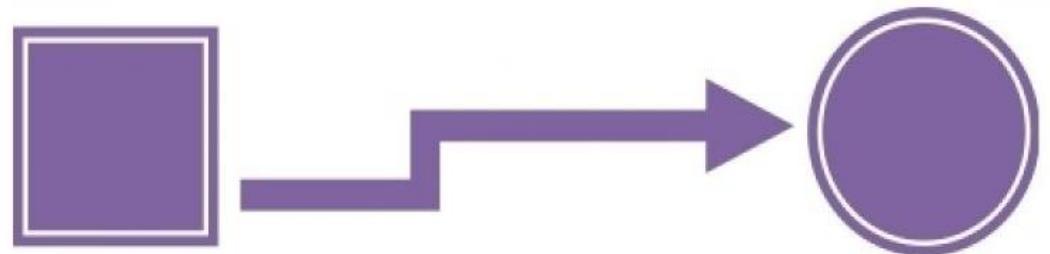
We still often suffer from technical problems like:

- Inadequate screen resolution
- Broken audio
- Grainy video
- Poor print quality
- Anything that results in the receiver receiving something different than what was originally crafted.



# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem
- How precisely do the transmitted symbols convey the desired meaning?





# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

- The semantic problem occurs when we encode the message using inappropriate visualization types, or when the symbols chosen won't be understood by the person on the receiving end.
- For example, encoding a value using a circle's diameter rather than its area will skew the perceived proportions



# Communication problems

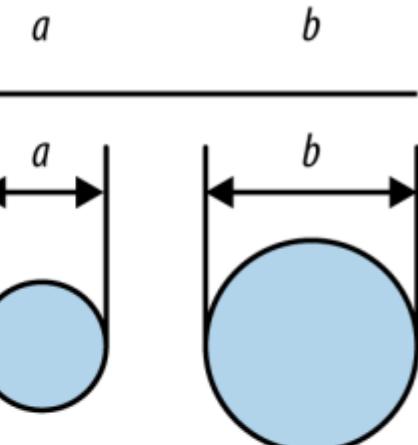
- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

**Values to encode:**

**Wrong:**

diameters proportional  
to values

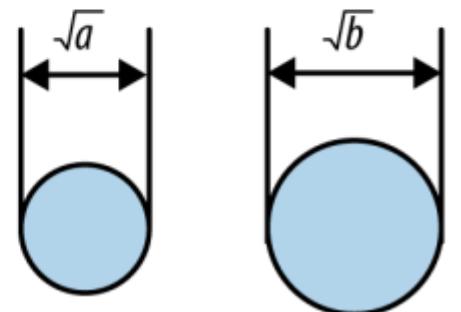
areas proportional  
to values<sup>2</sup>



**Right:**

diameters proportional  
to square root of values

areas proportional  
to values

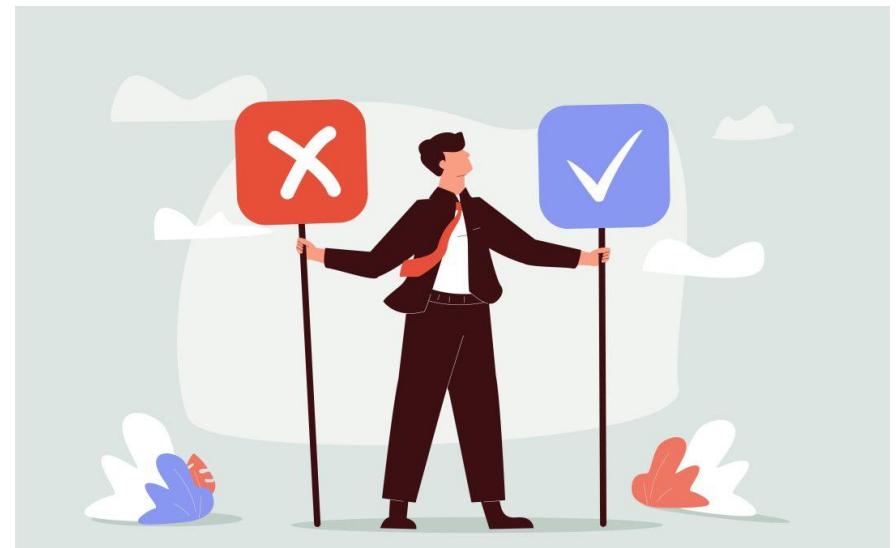




# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

- How effectively does the received meaning affect conduct in the desired way?





# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem





# Communication problems

- We can communicate data well, it's helpful to consider the types of communication problems that Shannon and Weaver identified.
  1. Technical problem
  2. Semantic problem
  3. Effectiveness problem

- The effectiveness problem is the “so what?” problem, and it might be the most important.
- If everything falls into place, and the message is perfectly encoded, transmitted, decoded, and understood, but the recipient doesn't care, or doesn't take the desired action, then the communication ultimately failed.



# Communicating data

1

A step in the process

2

A model of communication

3

Communication problems

4

Principles of communicating data



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

**Six  
Principles**



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- It's important to note that “information” and the “message” are not equal.
- **Information** is the set of all possible messages that can be selected by the information source.
- **The message** is what was selected from this set to be communicated.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- Before you choose your message, however, it's critical to know your goal, which you can articulate by answering a few key questions below:
  - Who are you trying to communicate with? **(Target audience)**
  - What do you want them to know? **(Intended meaning)**
  - Why? What do you want them to do about it? **(Desired effect)**



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results



*Figure 1-4. Elements of the goal*



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- Presentation of a single number: 14.
- That was the single data point shared with a group of managers assembled to discuss customer service.
- The number of times a particular customer had been transferred between departments during a single call to a helpline.
- It motivated an entire organization to revamp the customer experience.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- Shannon and Weaver identified this problem: “if you overcrowd the capacity of the audience, you force a general and certain error and confusion.”
- In other words, if a data point doesn’t add to your message, then it detracts from it.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- The last and most important point about selecting data is that your message must be both ethical and based on sound epistemology.
- In other words: don't lie with statistics—we have enough of that to struggle with already.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- Once you've identified the data the next step is deciding how to encode the message.
- Encoding the data means converting the data values themselves into abstract graphical representations, like size or color or shape.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## What type of data do you have?

- Tableau's owner Jock Mackinlay has produced a helpful framework for identifying the order of effectiveness of different encoding variables based on the type of data being used.
- First, let's start with a description of the different types of data: **quantitative**, **ordinal**, and **nominal**.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Quantitative

*Precise numerical values,  
for example:*

[2.54, 3.22, 10.12]

[4,2,9]

[1.2e12, 2.3e-7, -5.2e-8]



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Ordinal

*Elements related by order,  
for example:*

[small, med, large]

[1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> place]

[Disagree, Neutral, Agree]



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Nominal

*Members of a group or class, for example:*

[cat, dog, horse]  
[cash, credit, debit]  
[male, female]



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

What are the most effective types of visualizations for your data type?

- Once you've identified what data type or types you will need to get your point across, you need to decide what variables you will use to encode the data (see Figure 1~6).



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

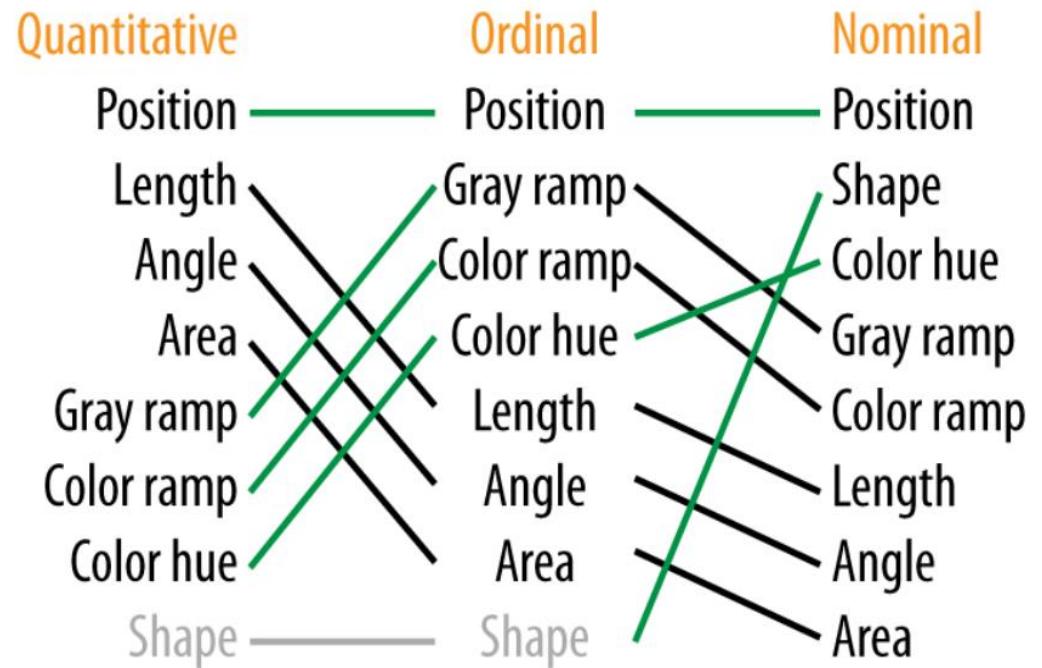


Figure 1-6. Effectiveness of data encoding



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

A few points are immediately obvious:

- Position is the most effective form of encoding for all data types.
- Length, angle, and area decrease in effectiveness from quantitative to ordinal to nominal.
- Color hue increases in effectiveness from quantitative to ordinal to nominal.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Few aesthetic elements:

- Poor color schemes
- Distracting fonts
- Many different fonts
- Sloppy alignment
- Vertical or angled labels
- Dark background colors
- Thick borders or grid lines
- Useless images and clip art
- Lazily accepting most software defaults

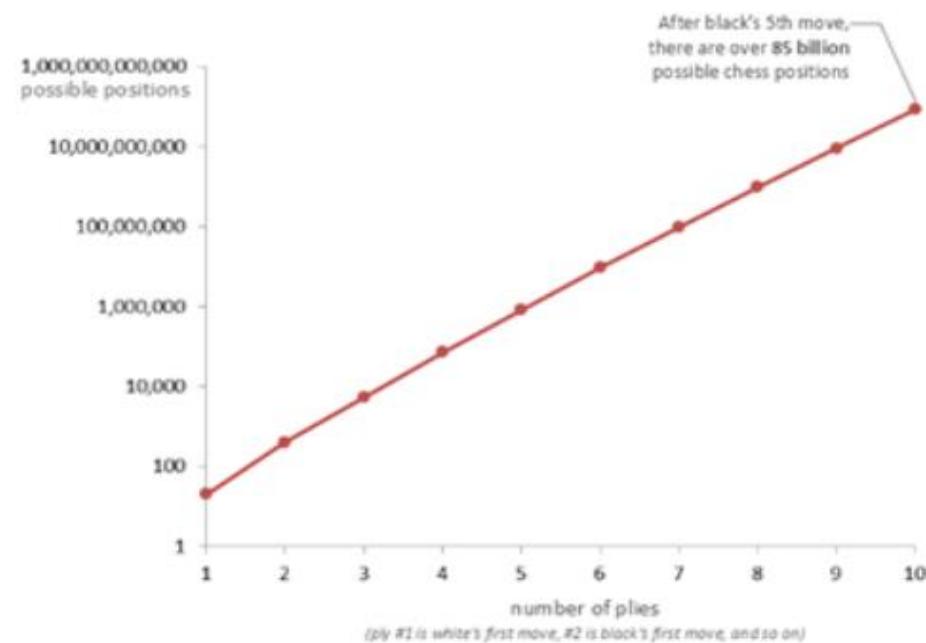
## Version A: Poor Aesthetics



## Version B: Aesthetics Improved

### Chess Positions Galore

François Labelle has calculated the number of distinct chess positions by ply, or move. The number increases dramatically with each successive move as the below line chart (with logarithmic y-axis) depicts. After just 5 moves each (or, by ply #10), the possibilities have reached staggering proportions



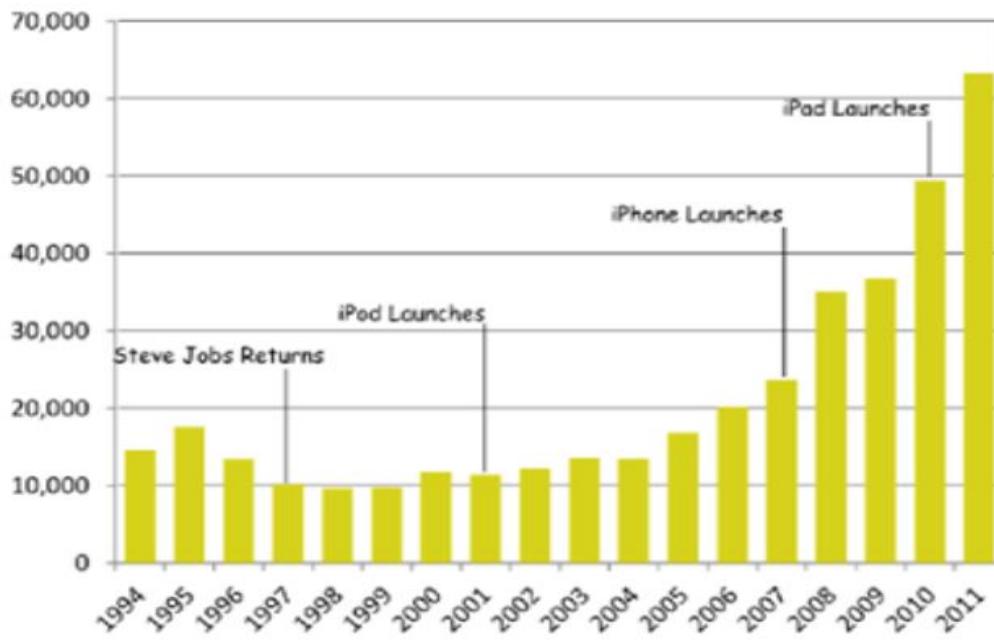
Source: <http://wismuth.com/chess/statistics-positions.html>

Figure 1-7. Two versions of the same line plot

## Version A: Poor Aesthetics

### Total Apple Employees

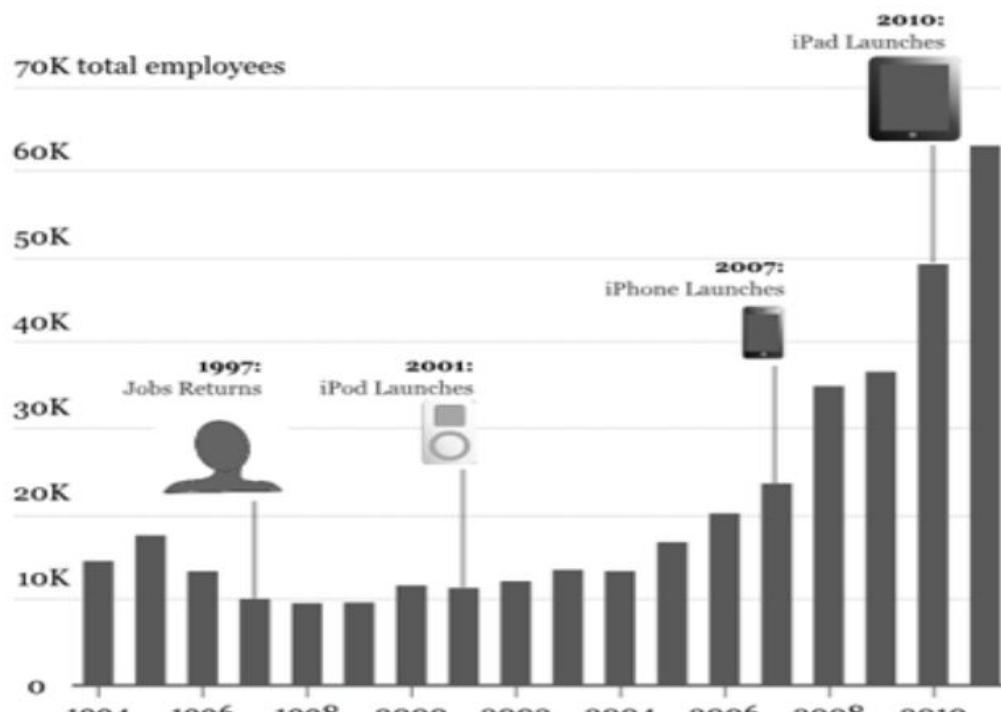
See how Steve Jobs's return to Apple was followed by innovative product launches and growth in employment



## Version B: Improved Aesthetics

### Apple, Inc. Employment History, 1994 - 2011

Between the return of Steve Jobs in 1997 and his death in 2011, Apple grew from 10,000 employees to over 63,000. A look at the timeline shows how key product launches fueled this growth:



Source: [www.sec.gov](http://www.sec.gov) | Ben Jones

Figure 1-8. Two versions of the column graph



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

When you communicate data, there are a few choices to make about how you will do it:

1. Standalone graphics or narrated?
2. Static, interactive, animated, or combined graphics?
3. If narrated: recorded, live, or both?
4. If live: remote, in person, or both?
5. In all cases: broadcast, directed, or both?



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

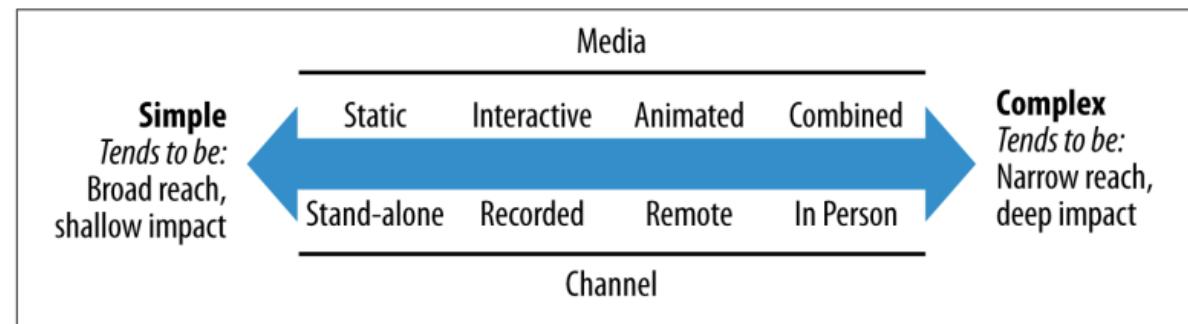


Figure 1-9. A spectrum of data communication types



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- It is a good habit in general to incorporate into your efforts feedback loops and checkpoints that help you measure whether you've achieved your intended results or not.
- This allows for course correction in the case of woefully unmet goals or fine-tuning in the case of slight miscues.



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

- There are a few questions to ask when you check the results.
- We'll call this the “RUI”:
  1. Reach
  2. Understanding
  3. Impact



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Reach

- Did the audience even receive your message at all? Who did and who didn't?



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Understanding

- Did the audience interpret the data message in the way you intended?



# Principles of communicating data

1. Know your goal
2. Use the right data
3. Select suitable visualizations
4. Design for aesthetics
5. Choose an effective medium and channel
6. Check the results

## Impact

- Did the audience react in the way you wanted them to react?



# Communicating data

1

A step in the process

2

A model of communication

3

Communication problems

4

Principles of communicating data



# DATA VISUALIZATION



# **UNIT - III (PART - II)**



# Introduction to Tableau

1

Using Tableau

2

Tableau products

3

Connecting to data



# Using Tableau

- Tableau software helps people communicate data through an innovation called VizQL, a visual query language.
- VizQL converts drag-and drop actions into data queries, allowing users to quickly find and share insights in their data.
- With Tableau, “data workers” first connect to data stored in files, cubes, databases, warehouses, Hadoop technologies, and even some cloud sources like Google Analytics.



# Using Tableau

- Data sources interact with the Tableau user interface to simultaneously query the data and view the results in charts, graphs, and maps that can be arranged together on dashboards.
- Communicate key insights, there are a variety of options depending on the product being used, from sending files to embedding interactive visualizations online to sharing via social media.



# Introduction to Tableau

1

Using Tableau

2

Tableau products

3

Connecting to data



# Tableau products

## 4 Types of Products

1. Tableau Desktop
2. Tableau Server
3. Tableau Online
4. Tableau Public



# Using Tableau

1. Tableau Desktop
2. Tableau Server
3. Tableau Online
4. Tableau Public

- A Windows application that comes in two editions Personal and Professional.
- Useful for analysts and business users.
- Personal allows connection to files and local saving only



# Using Tableau

1. Tableau Desktop
2. **Tableau Server**
3. Tableau Online
4. Tableau Public

- Best suited for enterprise-wide deployments, this is a business intelligence system for secure access to enterprise data and user interaction via web portals on a company intranet (requires Desktop Professional).



# Using Tableau

1. Tableau Desktop
2. Tableau Server
3. **Tableau Online**
4. Tableau Public

- A new hosted solution for storing and accessing data dashboards in the cloud (requires Desktop Professional), this is geared toward consultants and companies.



# Using Tableau

1. Tableau Desktop
2. Tableau Server
3. Tableau Online
4. Tableau Public

- The best option for journalists and bloggers, Tableau Public is a free application and visualization hosting service for sharing of publicly available data on the Web



# Using Tableau

- There are two other minor products that round out the offering:
  1. Tableau Public Premium
  2. Tableau Reader

- An annual subscription service that allows customers to prevent viewers of visualizations hosted on Tableau Public from downloading the workbook and accessing the underlying data



# Using Tableau

- There are two other minor products that round out the offering:
  1. Tableau Public Premium
  2. Tableau Reader

- A free Windows application that allows users to open saved Tableau workbook files (.twbx) and to view and interact with visualizations that have been created and saved locally with Tableau Desktop or downloaded from the Web via Tableau Public.



# Using Tableau

- There are two other minor products that round out the offering:
  1. Tableau Public Premium
  2. Tableau Reader
- Users of Tableau Reader cannot create new visualizations or change the design of existing ones.

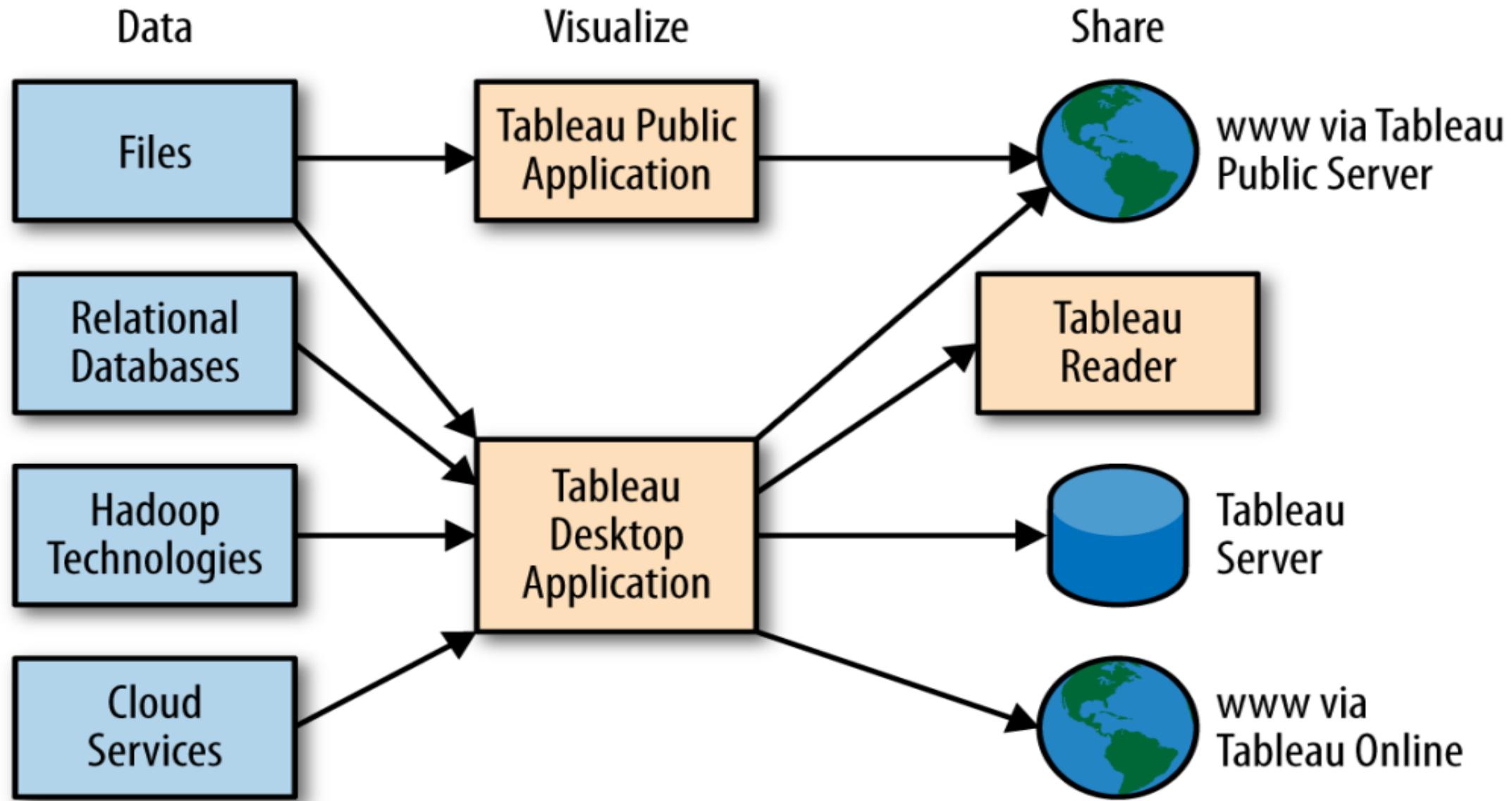


Figure 2-1. Tableau product diagram



# Introduction to Tableau

1

Using Tableau

2

Tableau products

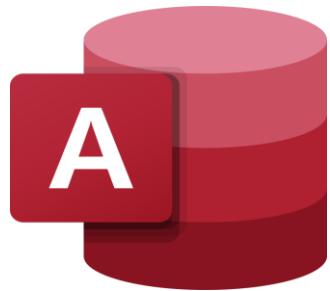
3

Connecting to data



# Connecting to data

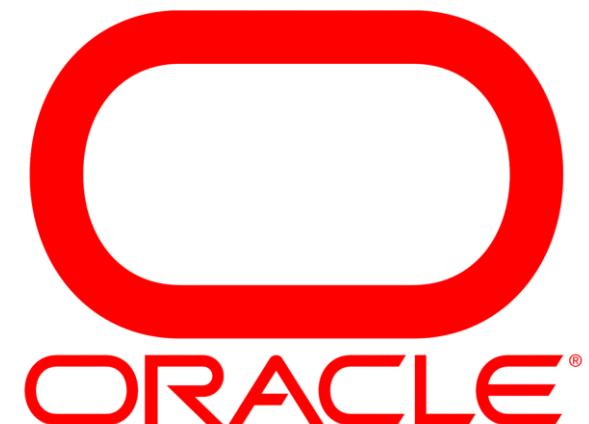
- Tableau Desktop Professional allows users to connect to data stored in the various data sources.
  1. Files
  2. Data Bases
  3. Hadoop Tech
  4. Cloud Sources





# Connecting to data

- Tableau Desktop Professional allows users to connect to data stored in the various data sources.
  1. Files
  2. Data Bases
  3. Hadoop Tech
  4. Cloud Sources





# Connecting to data

- Tableau Desktop Professional allows users to connect to data stored in the various data sources.
  1. Files
  2. Data Bases
  3. Hadoop Tech
  4. Cloud Sources





# Connecting to data

- Tableau Desktop Professional allows users to connect to data stored in the various data sources.
  1. Files
  2. Data Bases
  3. Hadoop Tech
  4. Cloud Sources





# Connecting to data

*The Tableau User Interface*



# Connecting to data

Every Tableau workbook contains

- **Sheets** are for creating individual visualizations.

- **Dashboards** are for combining sheets and other objects like images, text, and web pages on the same canvas, and adding interactions between them such as filtering and highlighting.



## Connect

Search for Data

Tableau Server

## To a File

Microsoft Excel

Text file

JSON file

Microsoft Access

PDF file

Spatial file

Statistical file

More...

## To a Server

Microsoft SQL Server

MySQL

Oracle

Amazon Redshift

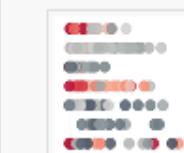
More...

## Saved Data Sources &gt;

## Open

Open a Workbook

## Accelerators



Superstore



World Indicators



Salesforce Pipeli...



Call Center Anal...



Purchasing

More Accelerators

## Discover

Meet Tableau

Get Started

Tour the Tableau Environment

Connect to and Prepare Data

Learn more...

Resources

Get Tableau Prep

Tableau Blueprint Assessment

Tableau Community Forums

Tableau Accelerators

Blog - Read latest post

File Data Server Help

- New Ctrl+N
- Open... Ctrl+O
- Hide Start Page Ctrl+2
- Paste Ctrl+V
- Repository Location...
- Exit

### To a File

- Microsoft Excel
- Text file
- JSON file
- Microsoft Access
- PDF file
- Spatial file
- Statistical file
- More...

### To a Server

- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More...

### Saved Data Sources

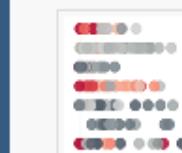
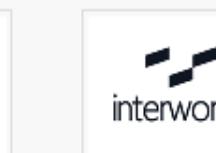
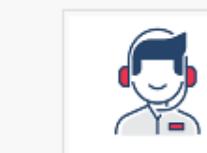
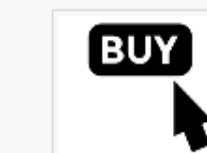
## Open

[Open a Workbook](#)

## Discover

[Meet Tableau](#)[Get Started](#)[Tour the Tableau Environment](#)[Connect to and Prepare Data](#)[Learn more...](#) [Resources](#)[Get Tableau Prep](#)[Tableau Blueprint Assessment](#)[Tableau Community Forums](#)[Tableau Accelerators](#)[Blog - Read latest post](#)

### Accelerators

[Superstore](#)[World Indicators](#)[Salesforce Pipeli...](#)[Call Center Anal...](#)[Purchasing](#)[More Accelerators](#)

Standard

Show Me

Data Analytics < Connect to Data

Tables

Pages Columns Rows

Filters

Marks

- Automatic
- Color
- Size
- Text
- Detail
- Tooltip

Sheet 1

Drop field here

Drop field here

Drop field here

Connect to data

Data Source Sheet 1

The screenshot illustrates the Tableau user interface for a new sheet. The interface is organized into several panels:

- Data Pane (Left):** Shows the "Data" section with a connection to "Sample - World Bank Indicators (Excel)". It lists dimensions: "Country / Region", "Date (year)", "Region", "Subregion", and "Measure Names". Below this is the "Measures" section containing numerous fields starting with "# B:" and "# F:", such as "ATMs (per 100k adults)", "Banks (per 100k adults)", "Biz tax rate (% profits)", etc.
- Marks Card (Left):** Located in the top left of the workspace, it shows the current mark type is "Automatic" and provides options for "Color", "Size", "Text", "Detail", and "Tooltip".
- Columns, Rows, and Pages (Top Left):** These are the primary shelf components where data fields are dropped to build the visualization.
- Filter Shelf (Top Left):** A vertical shelf for filtering data.
- Visualizations (Right):** A library of pre-designed charts and dashboards.
- Help and Tips (Bottom Right):** A panel titled "Select or drag data" with the instruction "Use the Shift or Ctrl key to select multiple fields".
- Numbered Callouts (Large Circles):**
  - 1:** Points to the Data pane.
  - 2:** Points to the Measures list.
  - 3:** Points to the "Select or drag data" tip.
  - 4:** Points to the Columns shelf.
  - 5:** Points to the workspace area.
  - 6:** Points to the Rows shelf.
  - 7:** Points to the Pages shelf.
  - 8:** Points to the Filter shelf.
  - 9:** Points to the bottom-left corner of the workspace.

**9** The Tableau user interface for a new Sheet



# Connecting to data

1. The list of data sources (can be more than one)
2. **Dimensions and Measures:** fields available to visualize in the selected data source
3. “Show Me” card: view applicable visualization types for selected fields
4. The **Columns and Rows shelves:** controls grouping headers (Dimensions) and axes (Measures)
5. The **Marks card:** control visualization encoding of color, size, label text, tooltip text, and shape.



# Connecting to data

6. **The Filters shelf:** filters visualizations by Dimensions or Measures.
7. **The Pages shelf:** filters the visualization by stepping or animating based on a particular field.
8. **The view itself:** this is the “canvas” where the data visualizations will appear.
9. **Sheets and Dashboard tabs:** show what has been created or create new Sheets or Dashboards
10. **The session tabs:** connect to data, show all tabs in a workbook, or see all workbooks for a user



← → ↕

Data

Analytics

Connect to Data

Tables

## New Dashboard

- Device Layouts
- Show Grid
- Grid Options...
- Format
- Copy Selected Dashboard Item
- Copy Image
- Export Image...
- Clear
- Show Title
- Actions... Ctrl+Shift+D
- Auto Update
- Run Update
- Add Phone Layouts to Existing Dashboards
- Add Phone Layouts to New Dashboards



Standard



Show Me

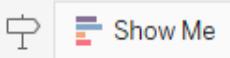
Drop field here

Drop field here



Connect to data





## Dashboard Layout &lt;

## Default

Phone

Device Preview

## Size

Desktop Browser (1000 x 8...

## Sheets

Sheet1

Drop sheets here



Connect to data

## Objects

Horizontal Container

Vertical Container

Text

Extension

Ask Data

Data Story

Image

Blank

Workflow

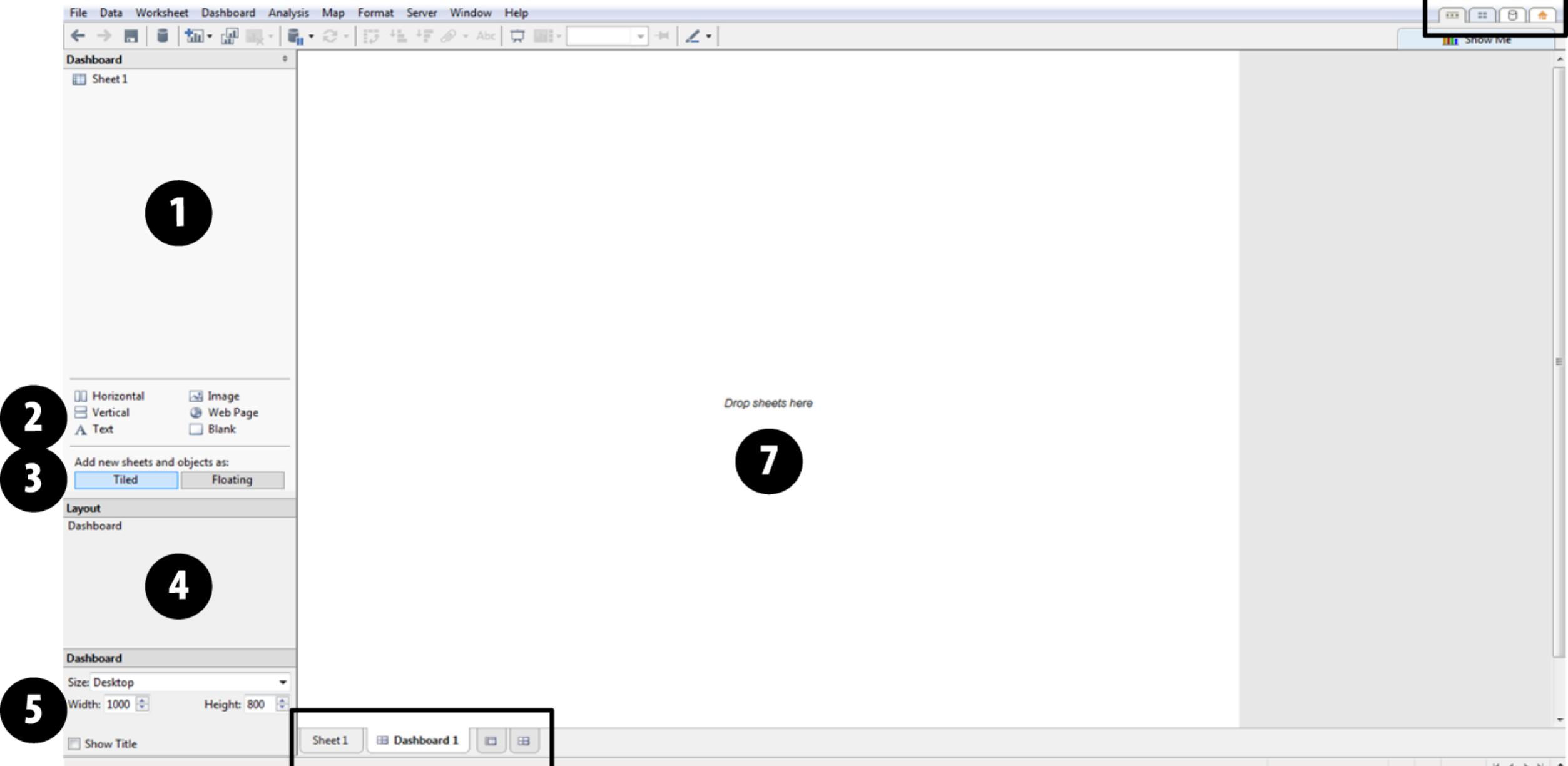
Web Page

Data Source

Sheet1

Dashboard 1





*The Tableau user interface for a new Dashboard*



# Connecting to data

1. The **list of Sheets** created in the current workbook
2. Dashboard **objects** to add (images, text, etc.)
3. **Tiled versus Floating object control** affects the object being dragged onto the dashboard
4. **Dashboard Layout outline** shows all sheets and objects included on a dashboard
5. **Dashboard size control** allows the users to specify the width and height of the dashboard.
6. **Sheets and Dashboard tabs:** show what has been created or create new Sheets or Dashboards
7. The **Dashboard** itself



# Connecting to data

## The tool bar:

- The components in both the Sheet view and the Dashboard view include a toolbar and menu items at the top, by default.

- The toolbar includes the all-important Undo (left arrow) and Redo (right arrow).
- Also included in the toolbar are controls to Save, Connect to Data, Sort, Group, Show Labels, Toggle to Presentation Mode, and change the Fit of the sheets, among a few other icons.



# Connecting to data

## Data types:

- When a user connects to a data source, Tableau automatically classifies each field as either a Dimension or a Measure.
- It's helpful to think of Dimensions as fields you can use to group or categorize your data;

- Dimensions can be further grouped into strings, dates, and geographic fields (which generate latitude and longitude Measures).
- Measures are fields you can do math with, like summing or averaging.
- Measures can be either discrete or continuous.



# Connecting to data

A	B	C	D	E	F
1	Borough	County	State	Population	Area sq-mi
2	MANHATTAN	New York	New York	1,619,090	23 2011
3	BRONX	Bronx	New York	1,408,473	42 2011
4	BROOKLYN	Kings	New York	2,565,635	71 2011
5	QUEENS	Queens	New York	2,272,771	109 2011
6	STATEN ISLAND	Richmond	New York	470,728	58 2011

*Figure 2-4. Data table showing population and area in New York boroughs*

	A	B	C	D	E	F
1	Borough	County	State	Population	Area sq-mi	Year
2	MANHATTAN	New York	New York	1,619,090	23	2011
3	BRONX	Bronx	New York	1,408,473	42	2011
4	BROOKLYN	Kings	New York	2,565,635	71	2011
5	QUEENS	Queens	New York	2,272,771	109	2011
6	STATEN ISLAND	Richmond	New York	470,728	58	2011

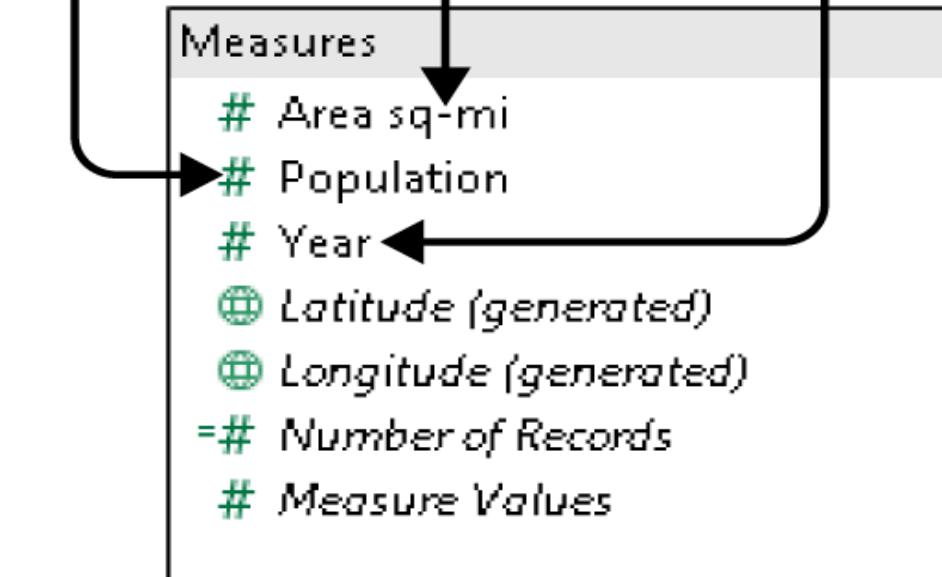
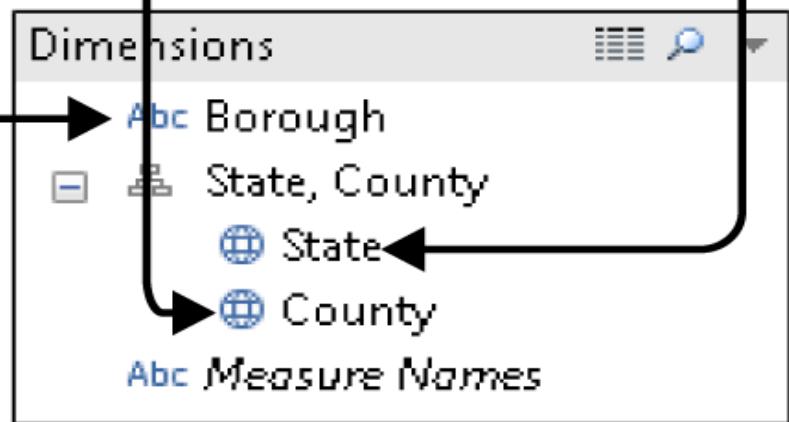


Figure 2-5. Tableau's default interpretation of the boroughs table



# Connecting to data

## Data types:

- By default, Tableau interprets Borough as a string (as indicated by the blue ABC to the left of the field name);
- County and State as geographic data types (as indicated by the blue globes next to these field names);
- Population, Area sq~mi, and Year as continuous measures (as indicated by the green # to the left of these field names).



# Connecting to data

- A few other fields appear in italics in the Dimensions and Measures areas.
- These fields aren't included in the original table but are rather generated by Tableau.

The generated dimensions include:

- **Measure Names** This field is a string type that provides a list of all the names of the fields in the Measures section
- In this case, Area sq~mi, Population, Year, and Number of Records.



# Connecting to data

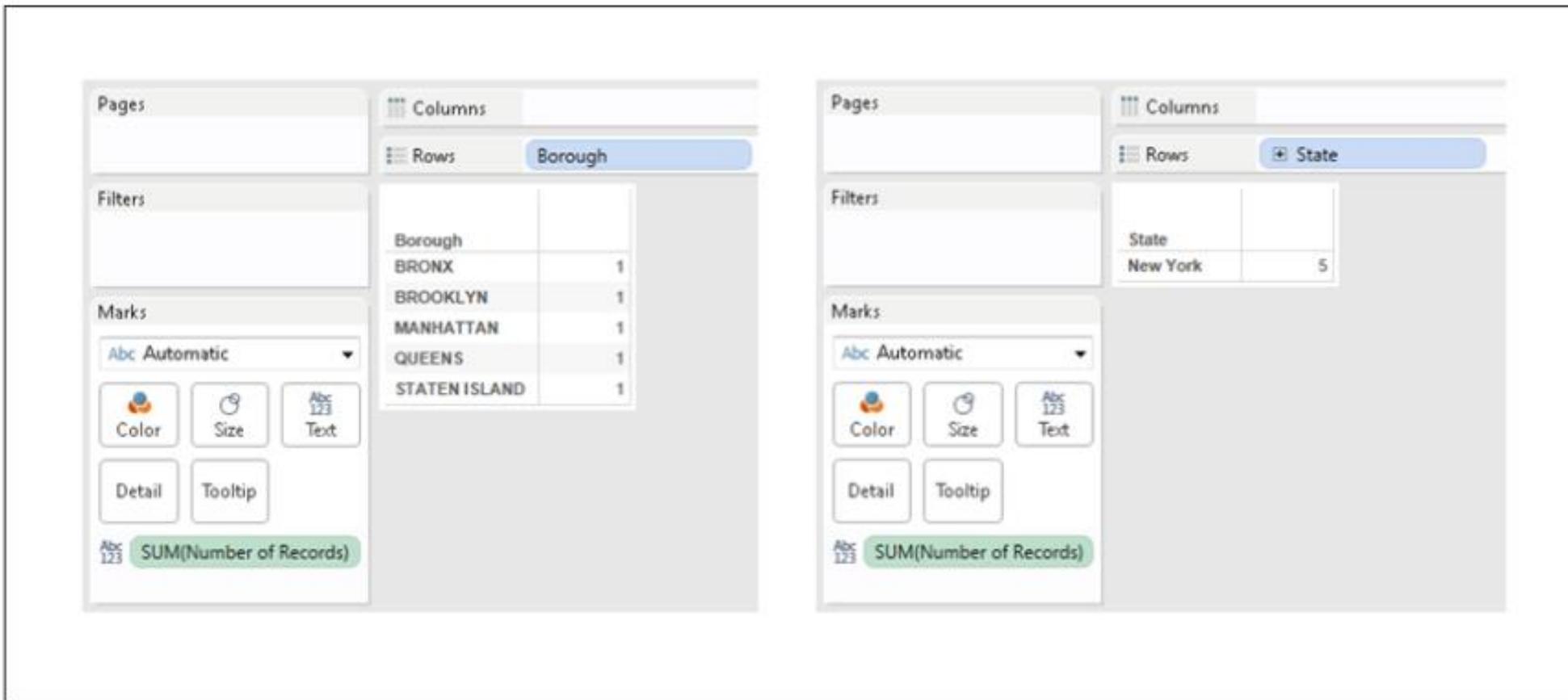
The generated measures include:

- **Measure Values:** This continuous numeric field contains all the numerical values of the Measures: every number included in the table, whether population, area, or year.

- **Number of Records** This numeric field is a count of the entries in the data source. This field is useful for showing how many of each dimension exists in the data set.
- **Latitude and Longitude** coordinates that correspond to the fields recognized as geographic data types (County and State).



# Connecting to data



*Figure 2-6. Illustration of the Number of Records field*

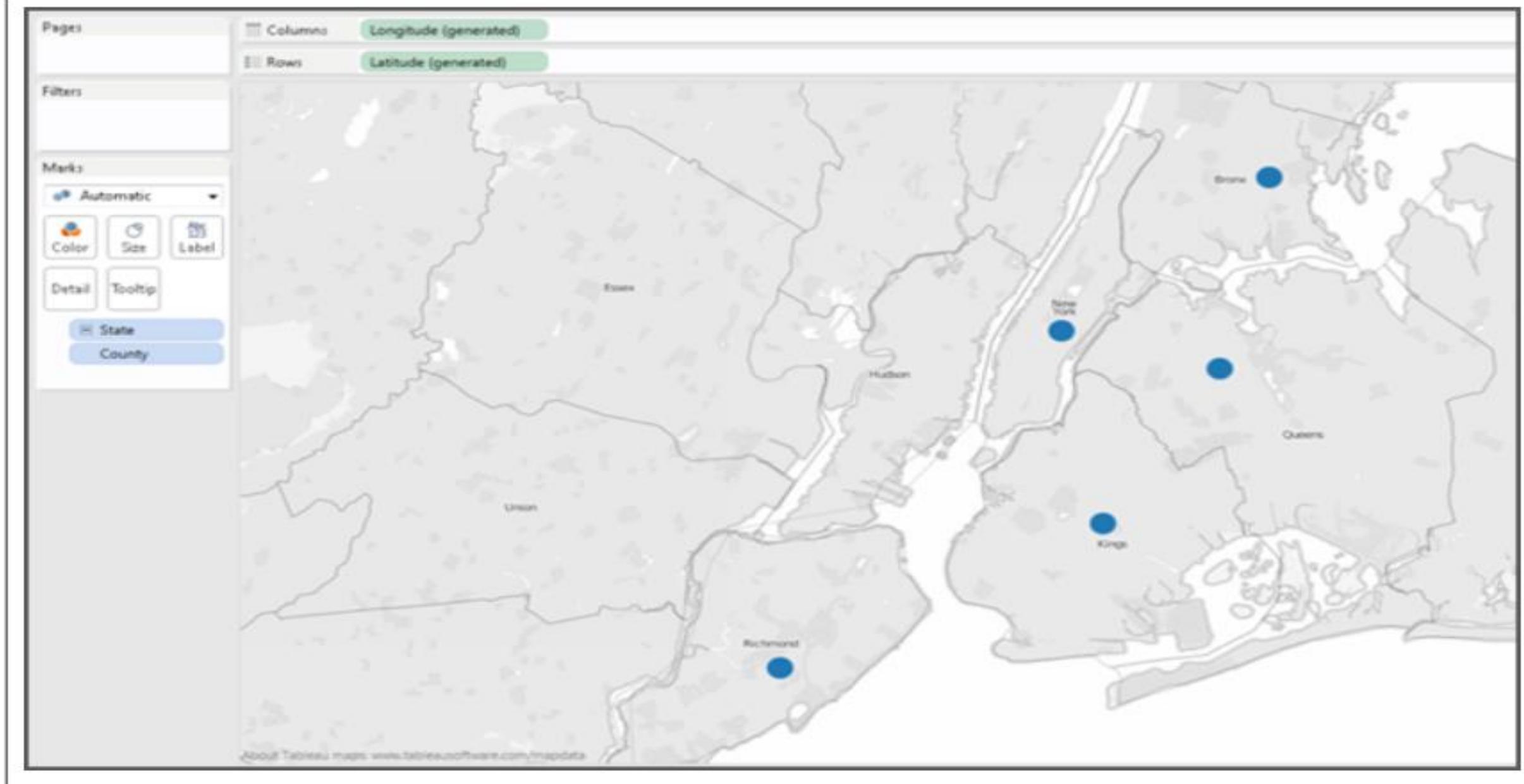


Figure 2-7. Symbol map for the geographic field “County”

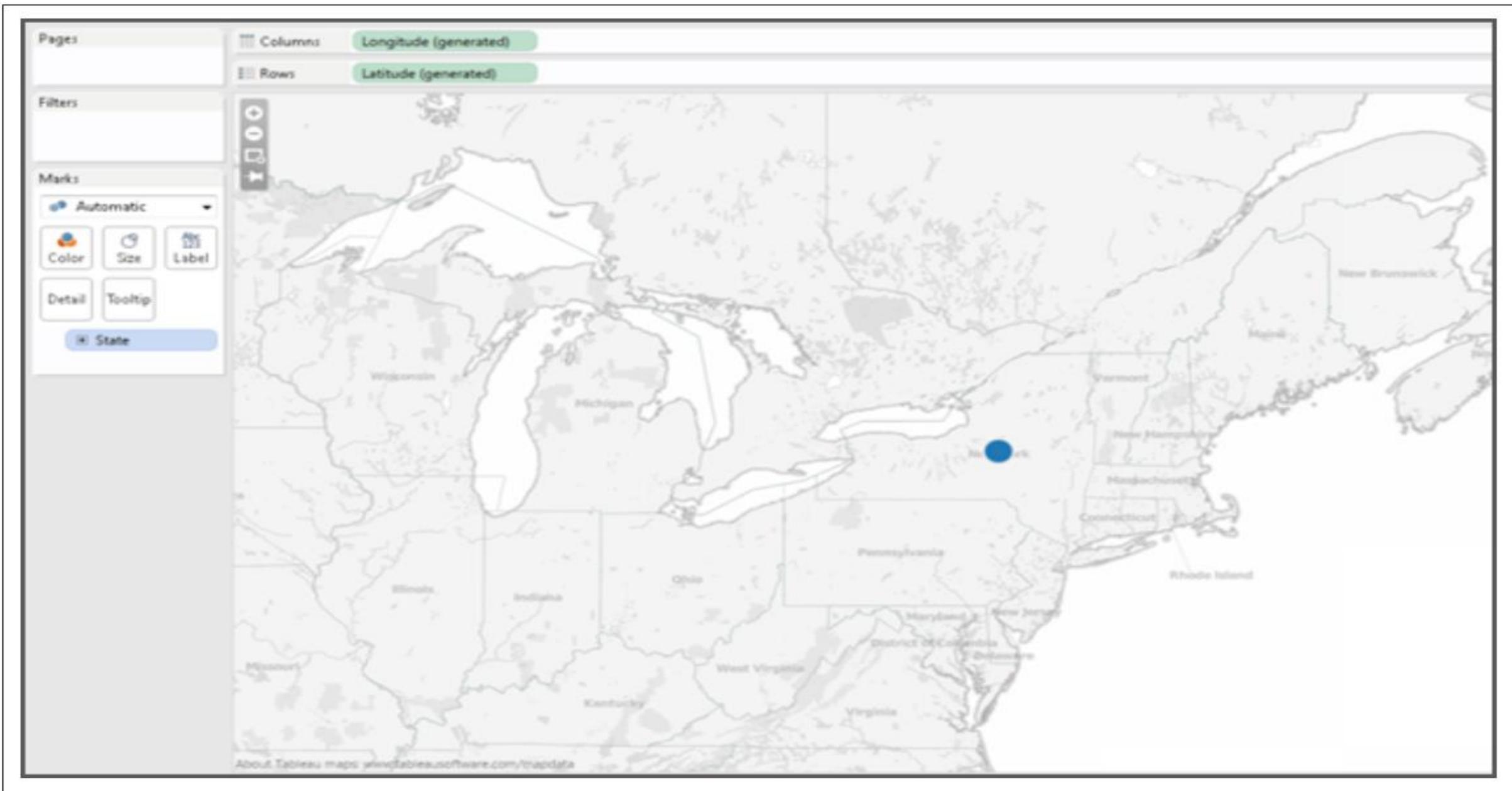


Figure 2-8. Symbol map for the geographic field “State”



# Connecting to data

## Changing data types:

- Tableau allows you to convert Dimensions to Measures and vice versa.
- Simply drag a field from one area of the screen to another, as shown in Figure 2~9

- The color of the # symbol next to the field name changes from green (for continuous) to blue (for discrete).
- This is another important distinction between data types.

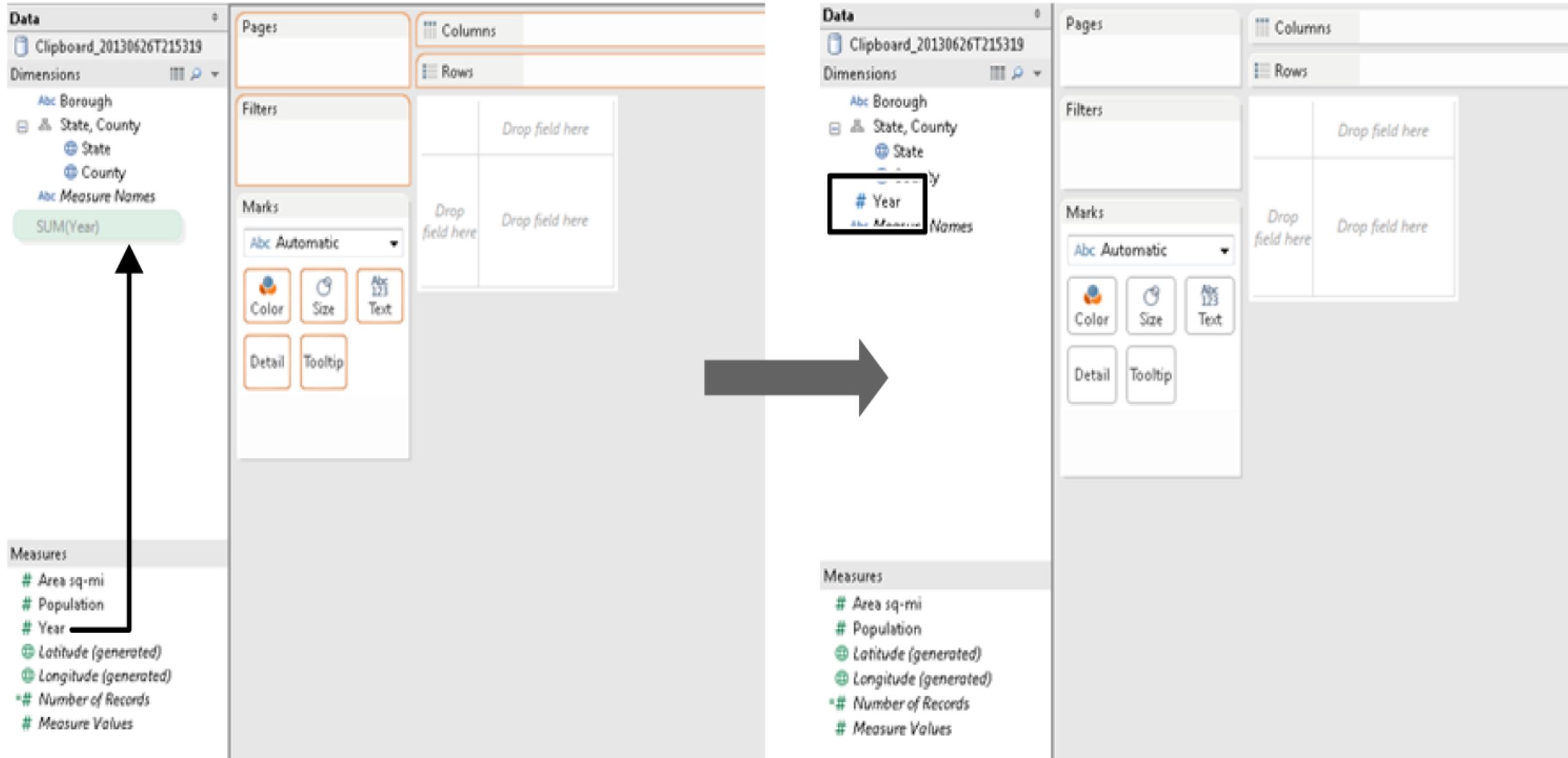


Figure 2-9. Changing a field from a Measure to a Dimension



# Connecting to data

- Tableau draws axes when continuous data types are used in visualizations (resulting in lines and scatterplots), and it creates headers when discrete data types are used (resulting in bars, columns, and tables).
- For example, consider the two visualizations in Figure 2-10, which show Area on the Columns shelf (the horizontal or x-axis) and Population on the Rows shelf (the vertical or y-axis).

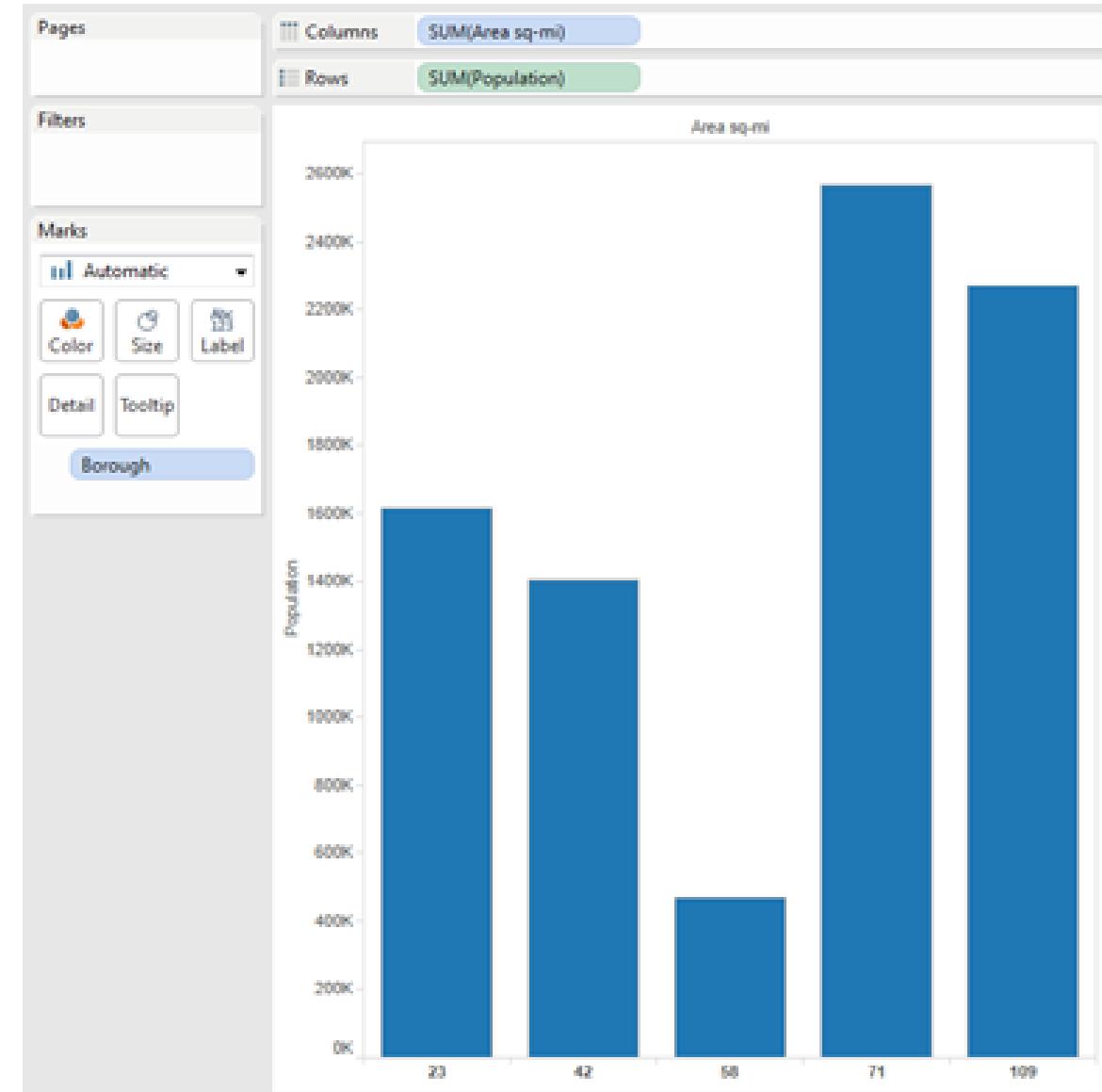
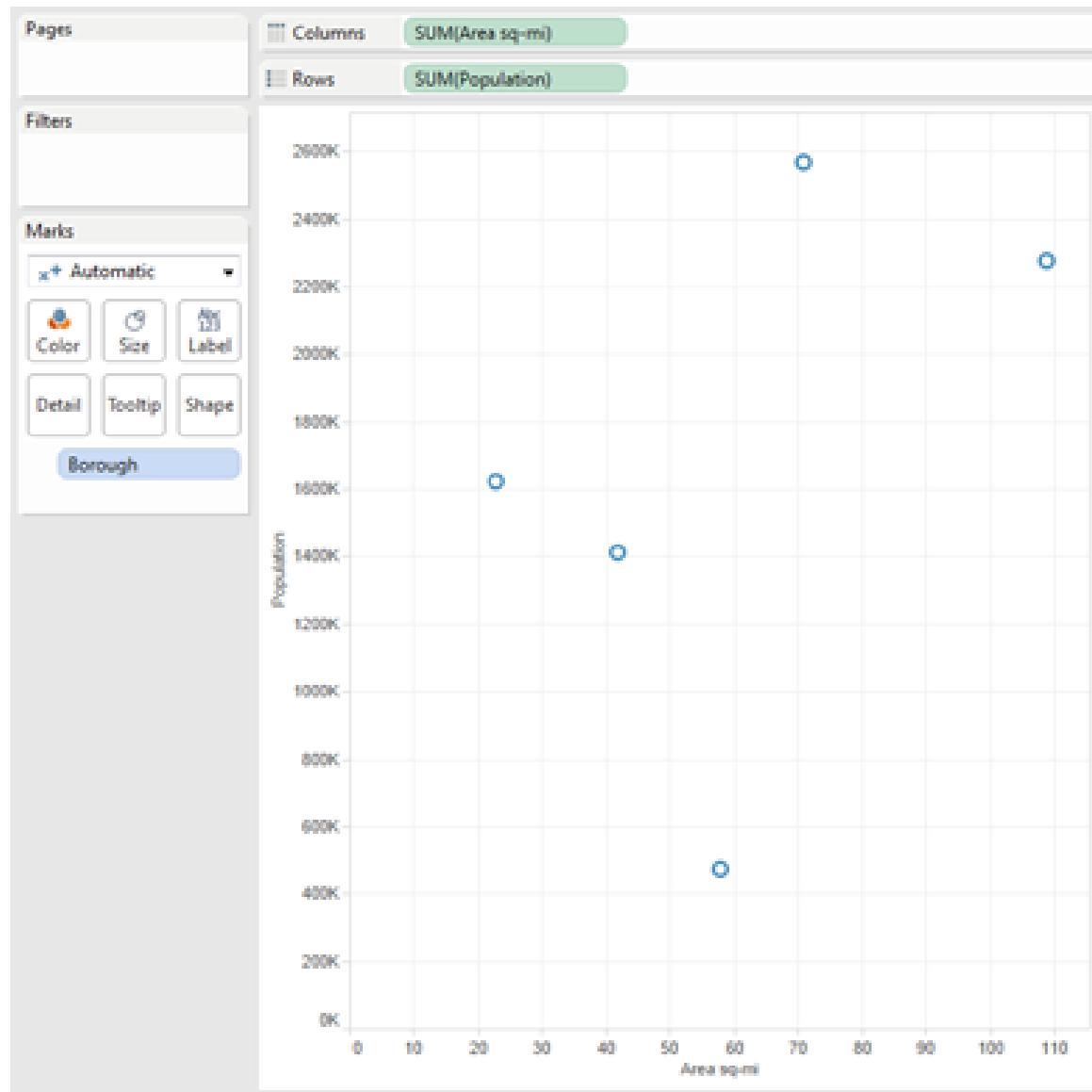


Figure 2-10. Area as a continuous data type (left); area as a discrete data type (right)

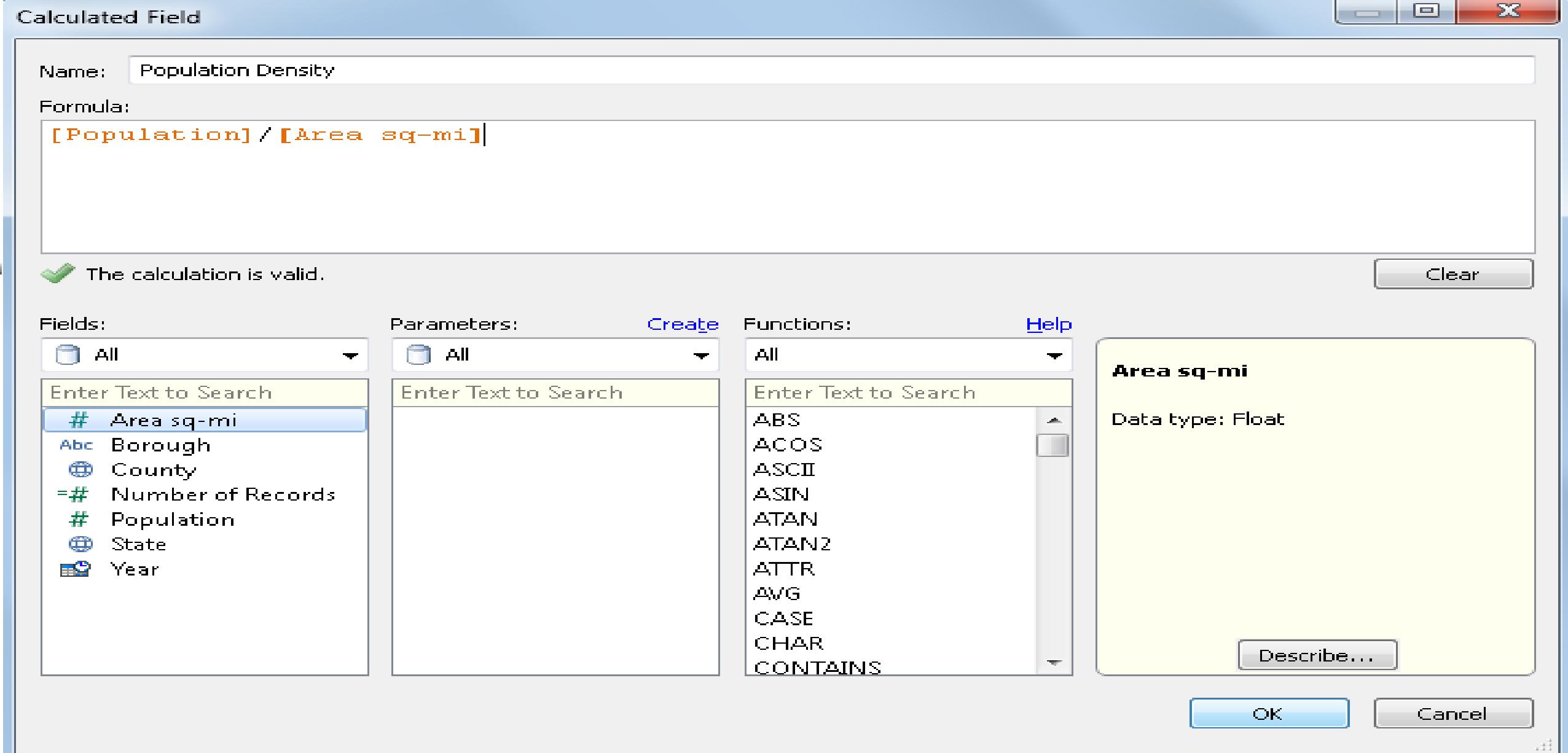


# Connecting to data

## Calculated fields:

- what if you wanted to figure out which borough in our example data table had the highest population density?

- The workbook already includes both population and area, so all you need to do is divide them to answer your question. With calculated fields, you can do just that.



*Figure 2-11. Creating a calculated field*

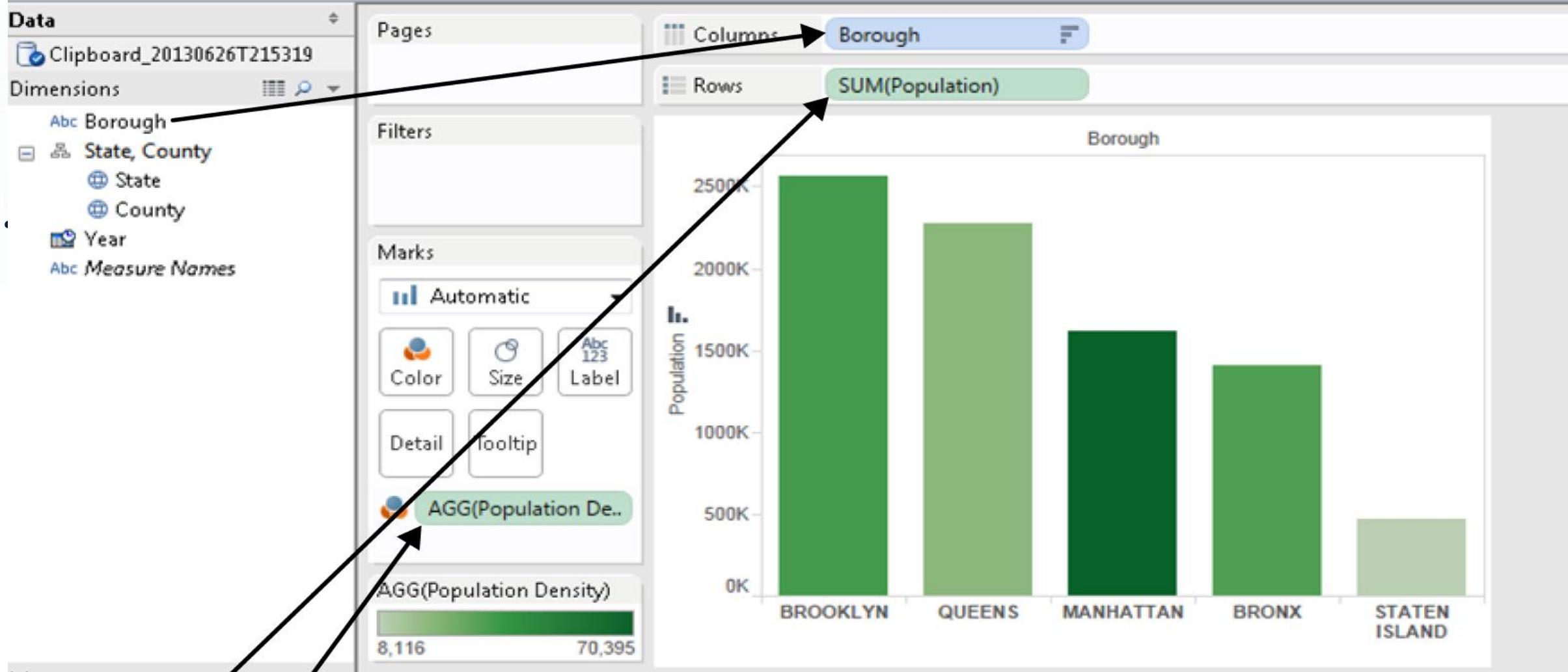


# Connecting to data

## Creating visualizations:

- Tableau features a simple drag-and-drop interface that allows the user to quickly explore different data visualization types.

- For example, dragging Borough onto the Columns shelf, Population onto the Rows shelf, and Population Density onto the Color shelf results in the column chart shown in Figure 2-12





# Connecting to data

## Creating visualizations:

- Another way to create a visualization in Tableau is to use the Show Me card.
- Show Me indicates which chart types are applicable based on the fields selected. Select multiple fields by holding the Control key.

- For example, if you select County, hold the Control key, and then select Population Density, the Show Me card will show a number of possible visualizations.
- Selecting the filled map icon in the top right results in the choropleth map of New York counties shown in Figure 2-13.

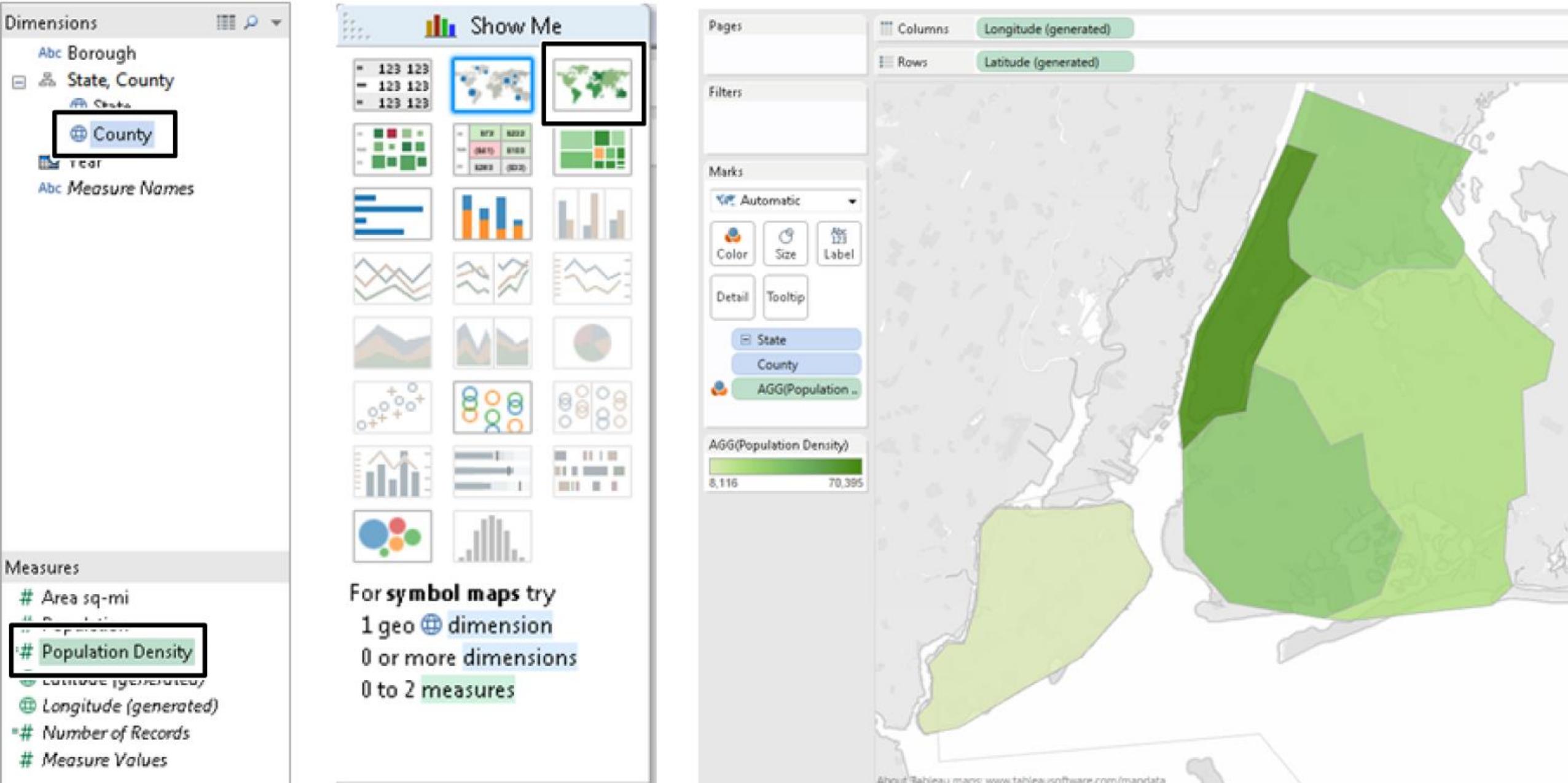


Figure 2-13. Creating a filled map in Tableau using Show Me



# DATA VISUALIZATION



# **UNIT - III (PART - II)**



# How much & How many

**How much** involves noncountable nouns like revenue or measurements:

- How much money did we earn this past quarter?
- How much carbon dioxide is in the atmosphere?

**How many** involves countable nouns like people or distinct events:

- How many customers do we have?
- How many goals did players score this season?



# How much and How many

1

Communicating how much

2

Communicating how many

3

Ratios and Rates



# Communicating how much

- let's consider comparisons involving quantities that are measured rather than counted.
- Ex: revenue, weight, distance, and time, among countless others.
- Your data source doesn't include a record for each dollar, pound, mile, or hour

A single record could be a transaction like a sale or a shipment:

- A sales order resulting in revenue of \$95
- An overnight shipment weighing 5.2 lbs.
- A flight covering 2,408 miles in 5 hours and 28 minutes



# Communicating how much

## Example:

- How much garbage does the City of New York Department of Sanitation (DSNY) collect from each borough?
- Well, each borough is further divided into community districts.

- A summary table of the total tonnage collected from each of the 59 community districts during September 2011

	A	B	C	D	E	F
1	Month	Borough	CommunityDistrict	RefuseTonsCollected	PaperTonsCollected	MGPTonsCollected
2	2011 / 09	Manhattan	1	1241.24	420.95	216.91
3	2011 / 09	Manhattan	2	2245.23	500.94	269.77
4	2011 / 09	Manhattan	3	3178.24	388.52	257.06
5	2011 / 09	Manhattan	4	2279.71	517.98	335.69
6	2011 / 09	Manhattan	5	1348.82	278.28	152.43
7	2011 / 09	Manhattan	6	3117.88	628.98	385.61
8	2011 / 09	Manhattan	7	5461.86	1232.56	617.15
9	2011 / 09	Manhattan	8	5904.28	1225.57	643.79
10	2011 / 09	Manhattan	9	2717.01	291.66	190.29
11	2011 / 09	Manhattan	10	2962.33	224.69	156.22
12	2011 / 09	Manhattan	11	2329.99	180.59	110.67
13	2011 / 09	Manhattan	12	5281.14	434.96	380.6
14	2011 / 09	Bronx	1	2052.58	91.27	76.03
15	2011 / 09	Bronx	2	1664.38	82.04	97.48
16	2011 / 09	Bronx	3	1917.67	80.88	78.65
17	2011 / 09	Bronx	4	4588.12	169.44	167.9
18	2011 / 09	Bronx	5	3735.12	164.09	207.24
19	2011 / 09	Bronx	6	2328.64	95.04	112.25
20	2011 / 09	Bronx	7	3926.07	237.94	244.82
21	2011 / 09	Bronx	8	2651.04	351.2	237.54
22	2011 / 09	Bronx	9	4510.36	200.64	235.91
23	2011 / 09	Bronx	10	2860.96	378.97	274.43
24	2011 / 09	Bronx	11	3307.65	333.6	251.35
25	2011 / 09	Bronx	12	4185.39	390.3	404.84
26	2011 / 09	Brooklyn	1	4819.68	535.61	372.21

*Figure 3-1. DSNY Collection Tonnages data table*



# Communicating how much

- The first question to ask of the data set is a rather simple one.
- How does the amount of garbage (in tons) that DSNY reportedly collected from each borough during September 2011 compare?

- We'll start by connecting Tableau to the data set by clicking Connect to Data and browsing to the file.
- Once you have Tableau connected to the DSNY Collection Weights data set, you will see Sheet 1 as a blank canvas with the data fields arranged, as shown in Figure 3~2.

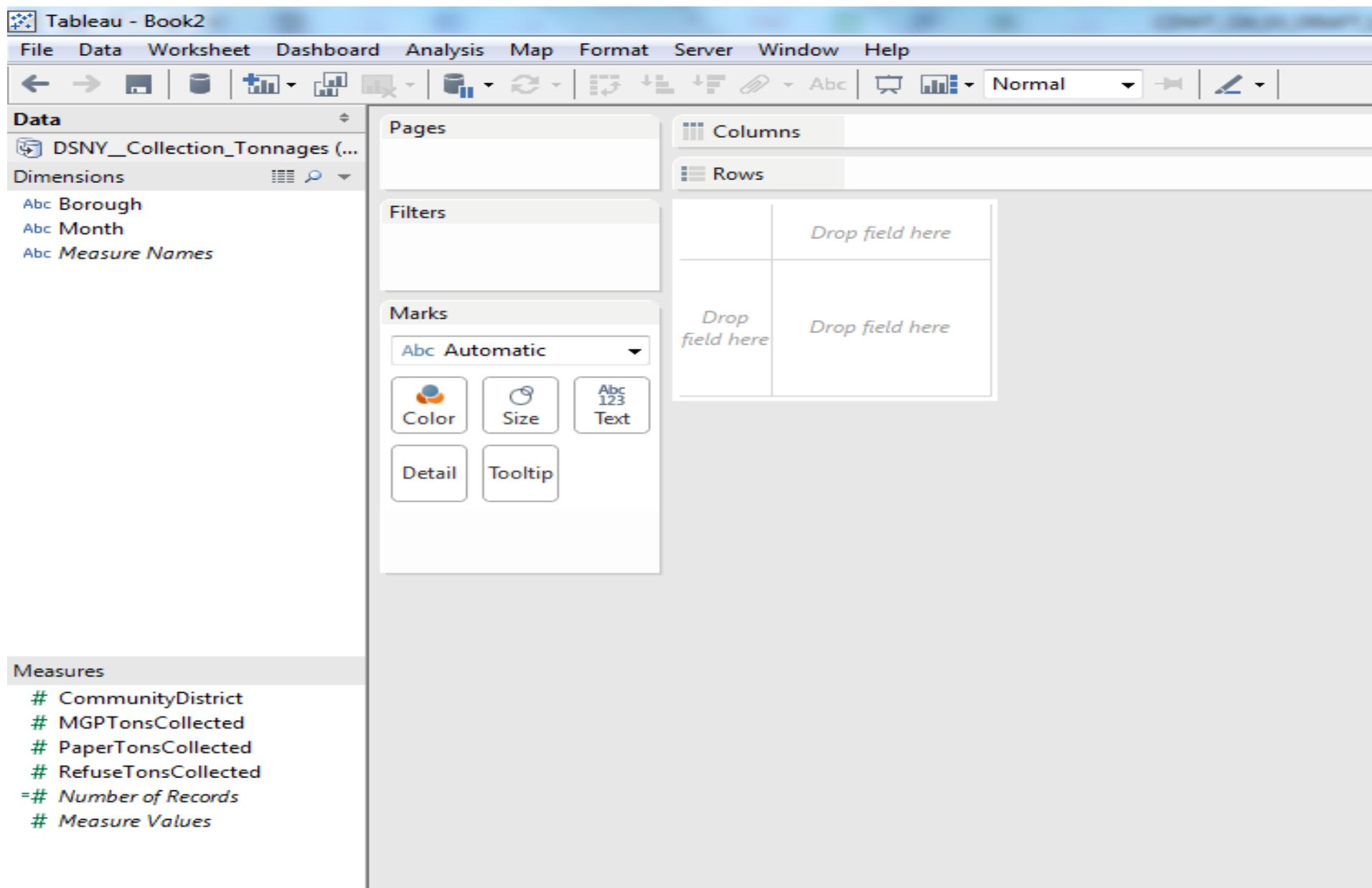


Figure 3-2. Connected to the DSNY Collection Tonnages data set

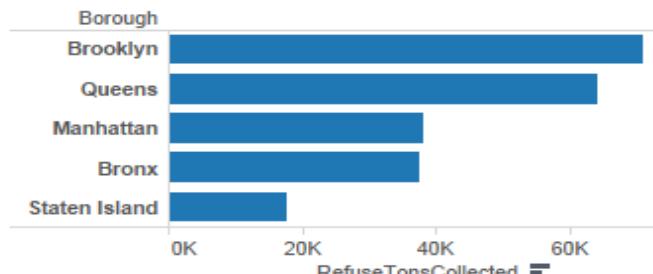


# Communicating how much

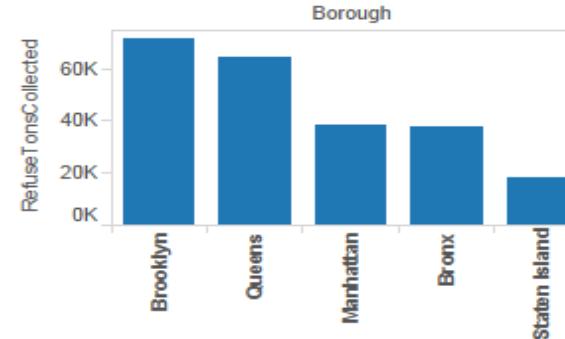
- Notice that Tableau has decided that Borough and Month are discrete (blue) string Dimensions.
- Community District and the three forms of garbage (MGP stands for “Metal, Glass, and Plastic”) collected are continuous (green) numerical Measures.
- To begin, click Borough, hold down the Control key, click Refuse Tons Collected, and then click Show Me in the upper-right corner of the screen.
- Tableau suggest chart type, based on the principles of data encoding effectiveness outlined in Figure 1~6.

## 11 Ways to Show Borough Garbage Collection Amounts

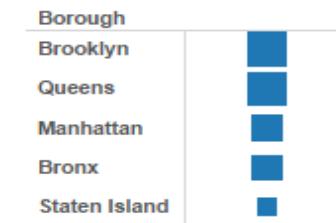
1. horizontal bars



2. vertical bars



3. heat map



RefuseTonsCollected

17,886
30,000
40,000
50,000
60,000
71,068

Borough
Brooklyn
Queens
Manhattan
Bronx
Staten Island

4. text table

Borough	RefuseTonsCollected
Brooklyn	71,068
Queens	64,365
Manhattan	38,068
Bronx	37,728
Staten Island	17,886

5. packed bubbles



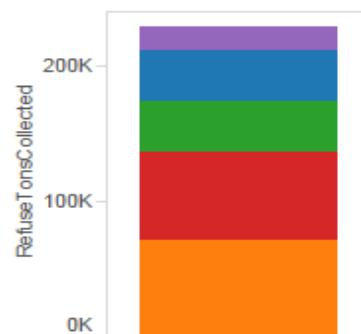
6. word cloud



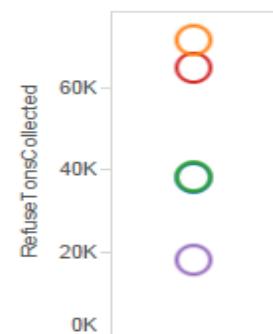
7. pie chart



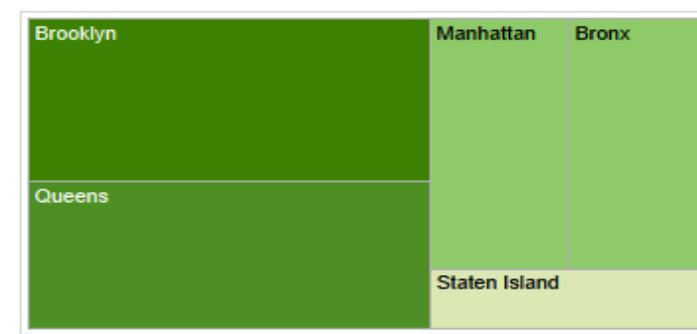
8. stacked bars



9. circle view



10. treemap



11. highlight table

Borough	RefuseTonsCollected
Brooklyn	71,068
Queens	64,365
Manhattan	38,068
Bronx	37,728
Staten Island	17,886



Figure 3-3. Eleven ways to show a simple quantitative (noncounting noun) comparison

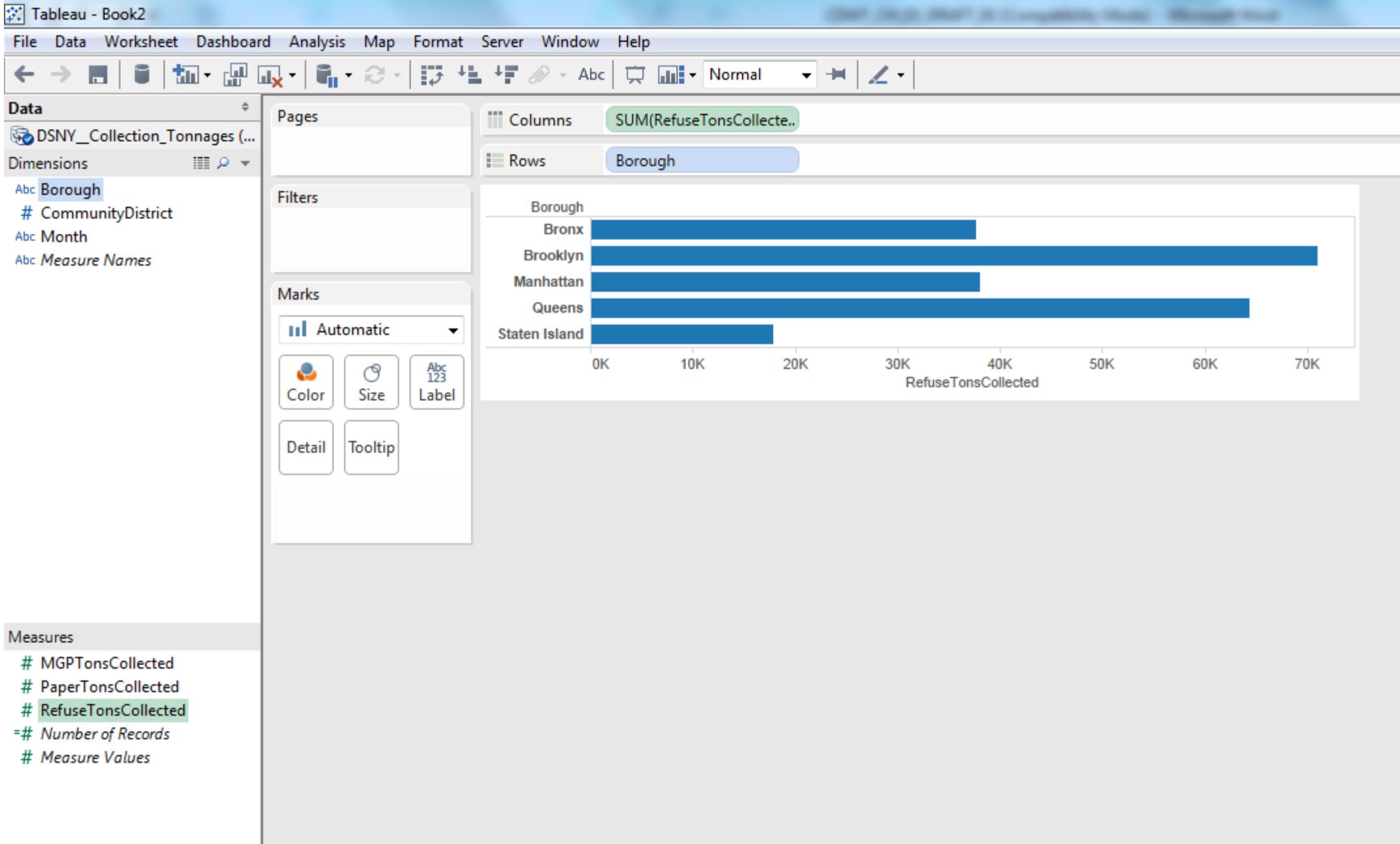
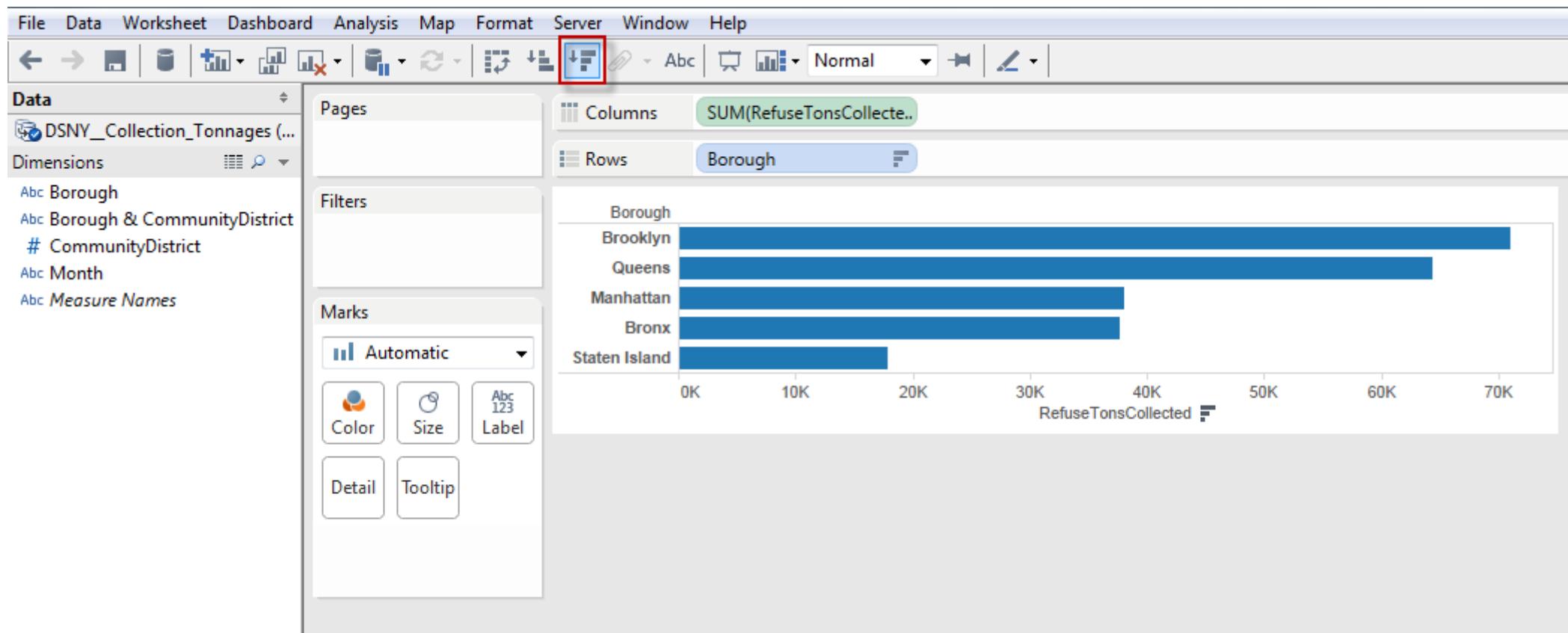


Figure 3-4. Horizontal bar chart of refuse tons collected



*Figure 3-5. Sorting using the toolbar icons*

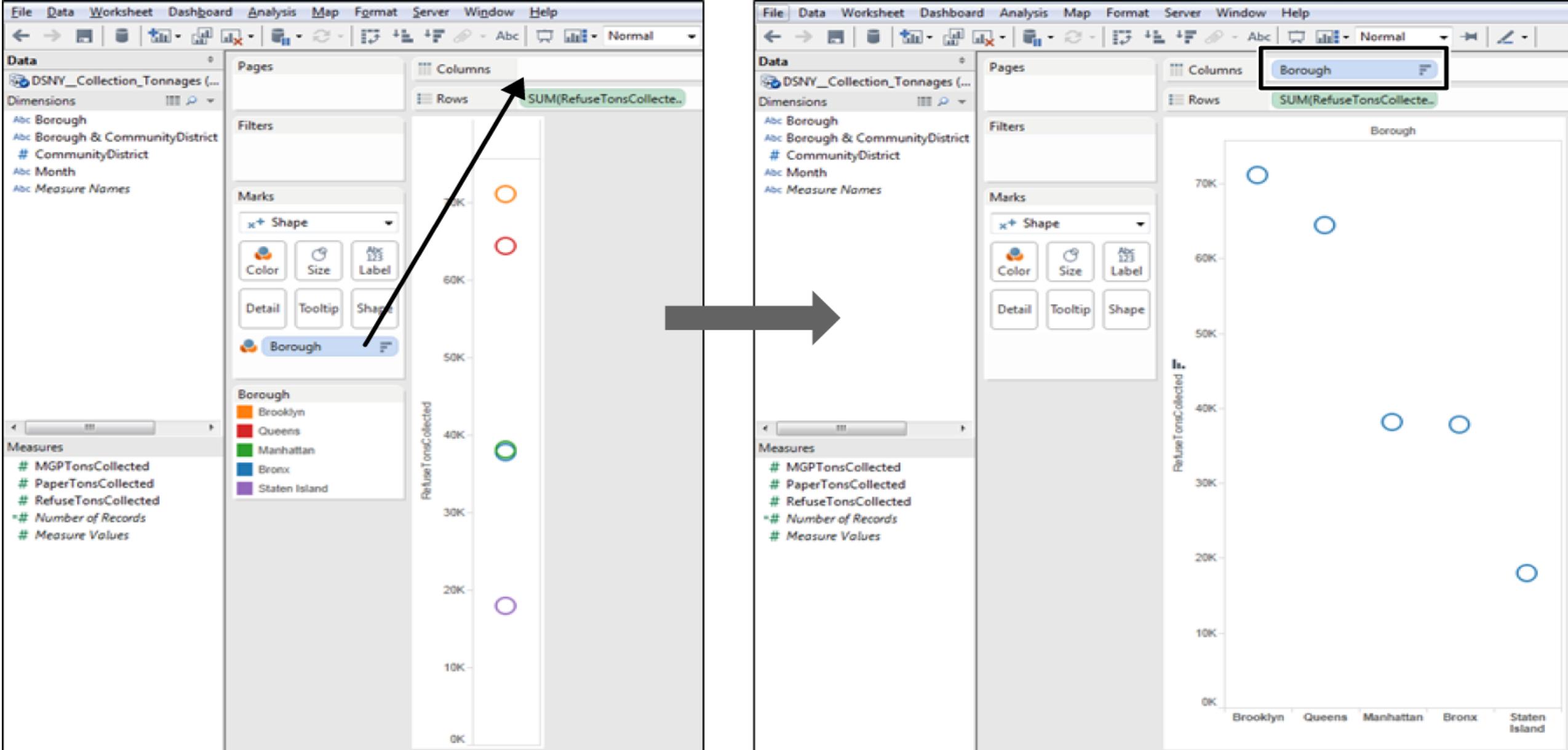


Figure 3-7. Resolving overlapping circles

Swap

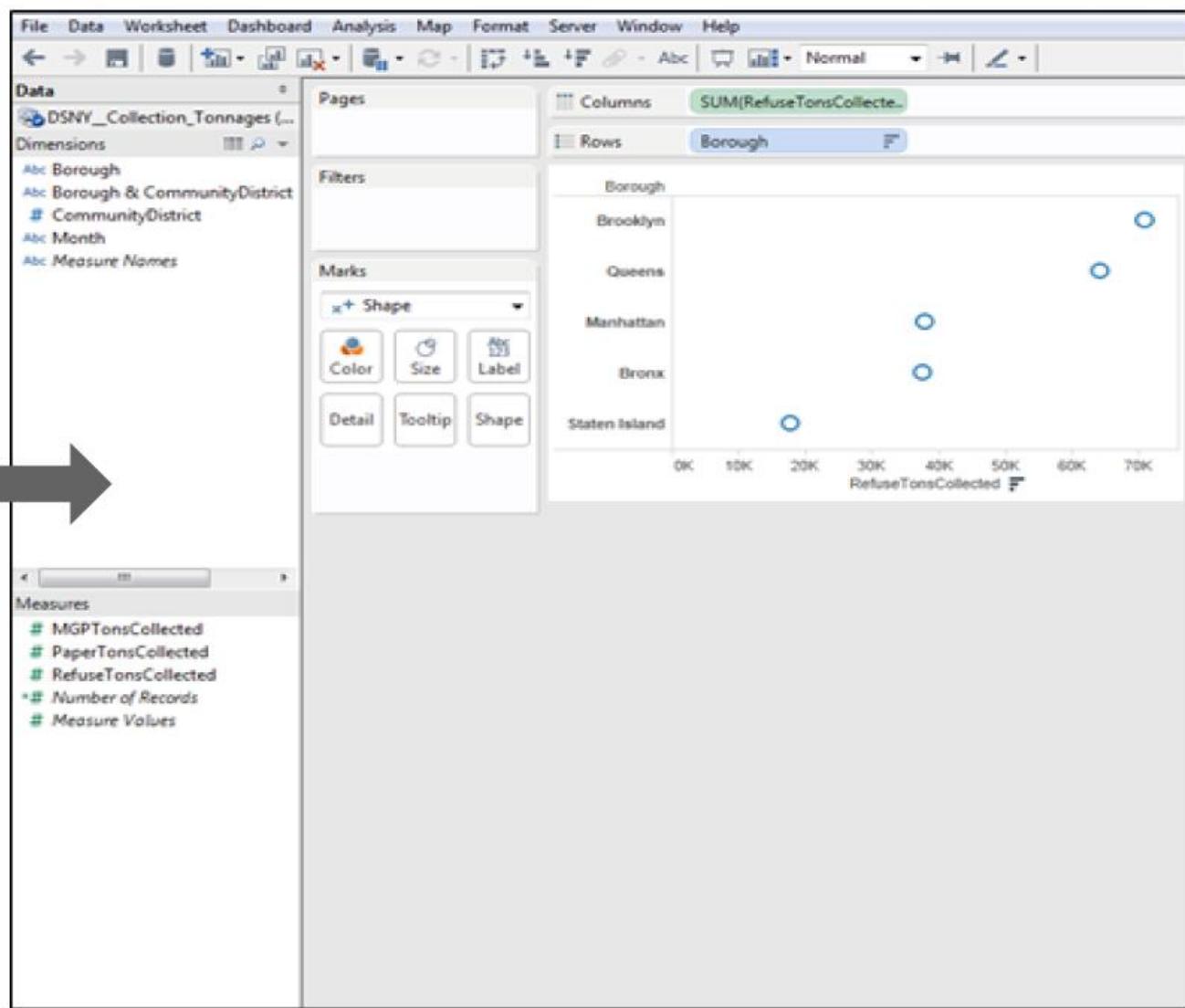
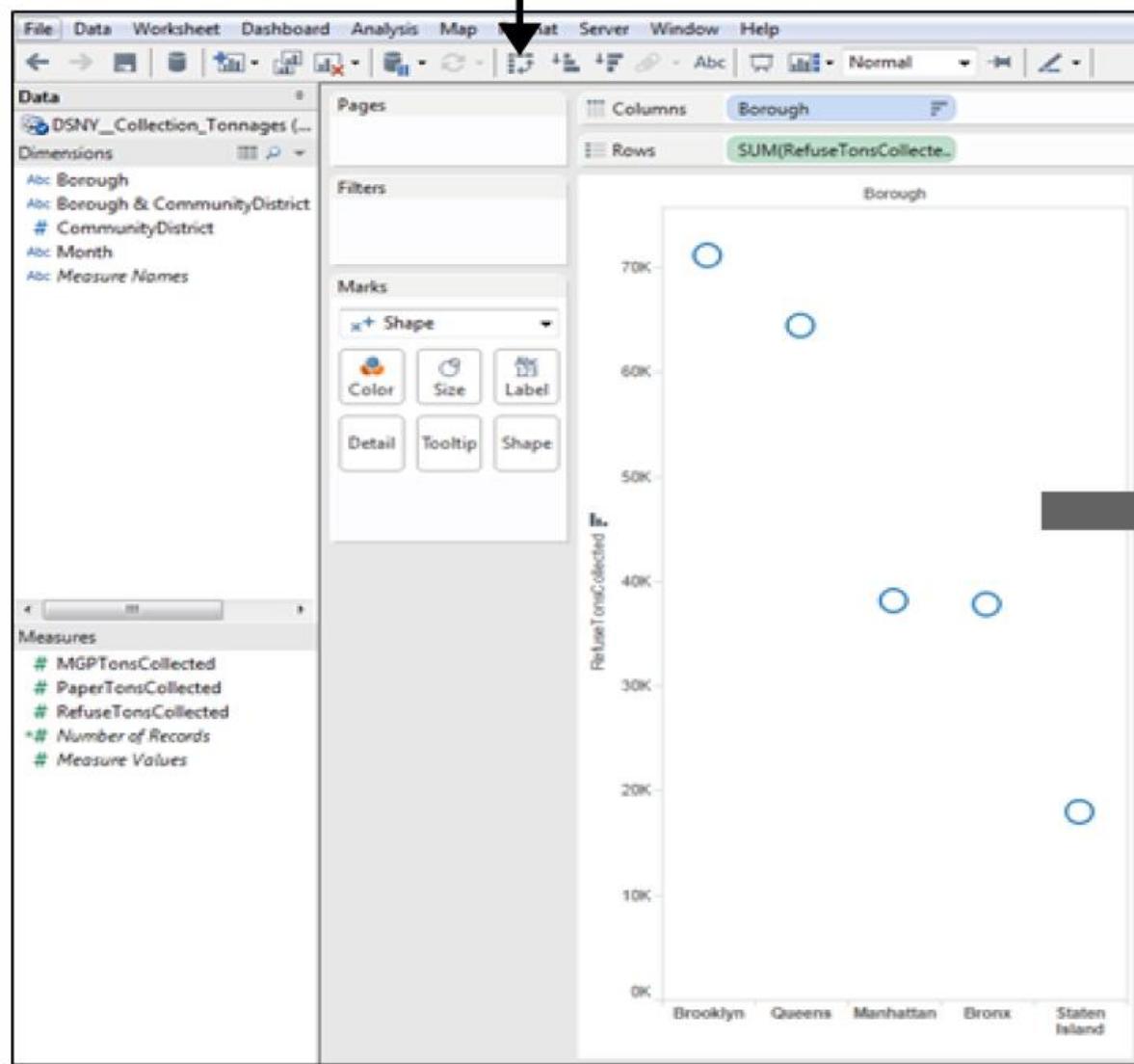


Figure 3-8. Using the Swap function to rotate view

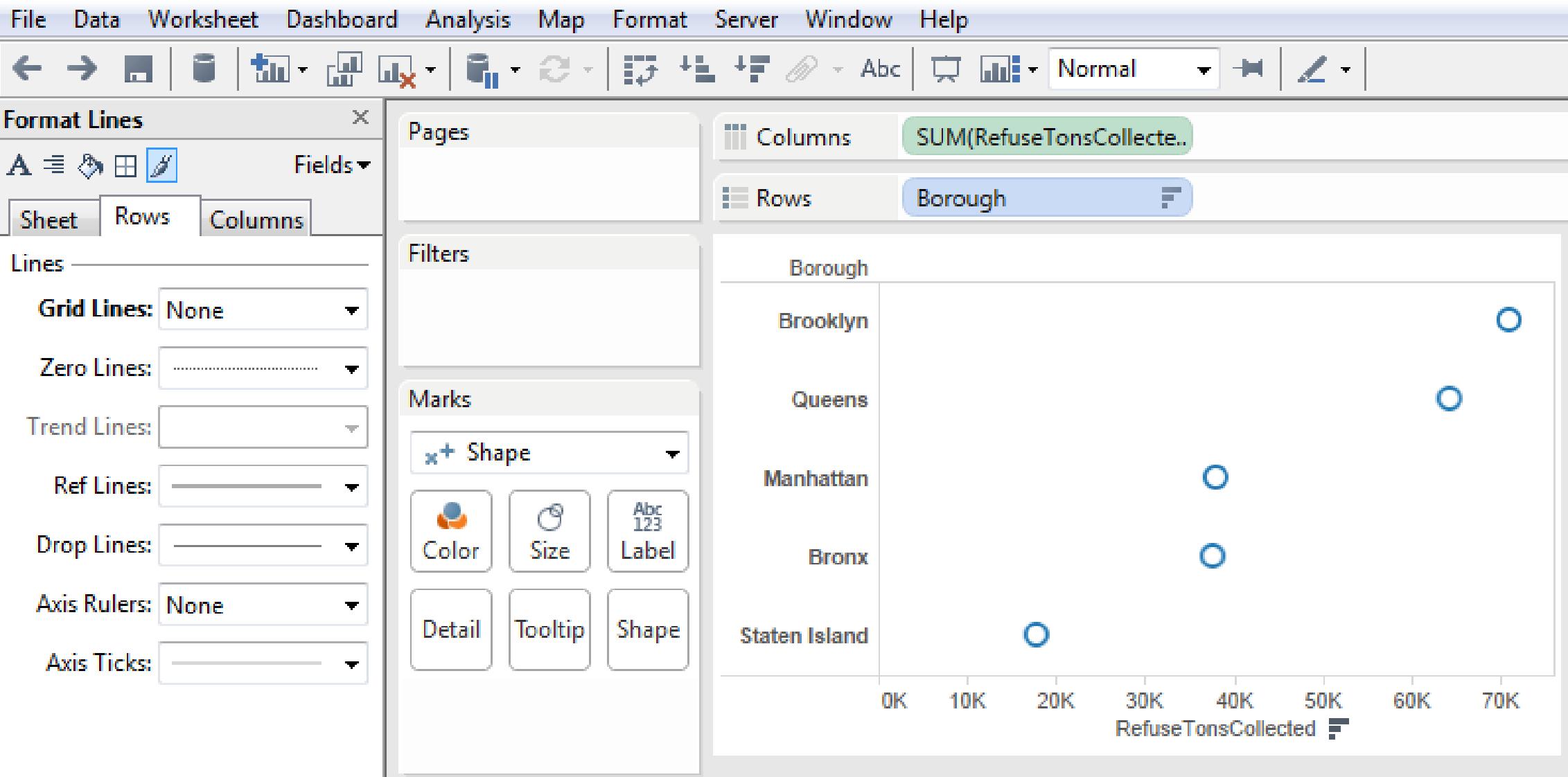


Figure 3-9. The Format panel

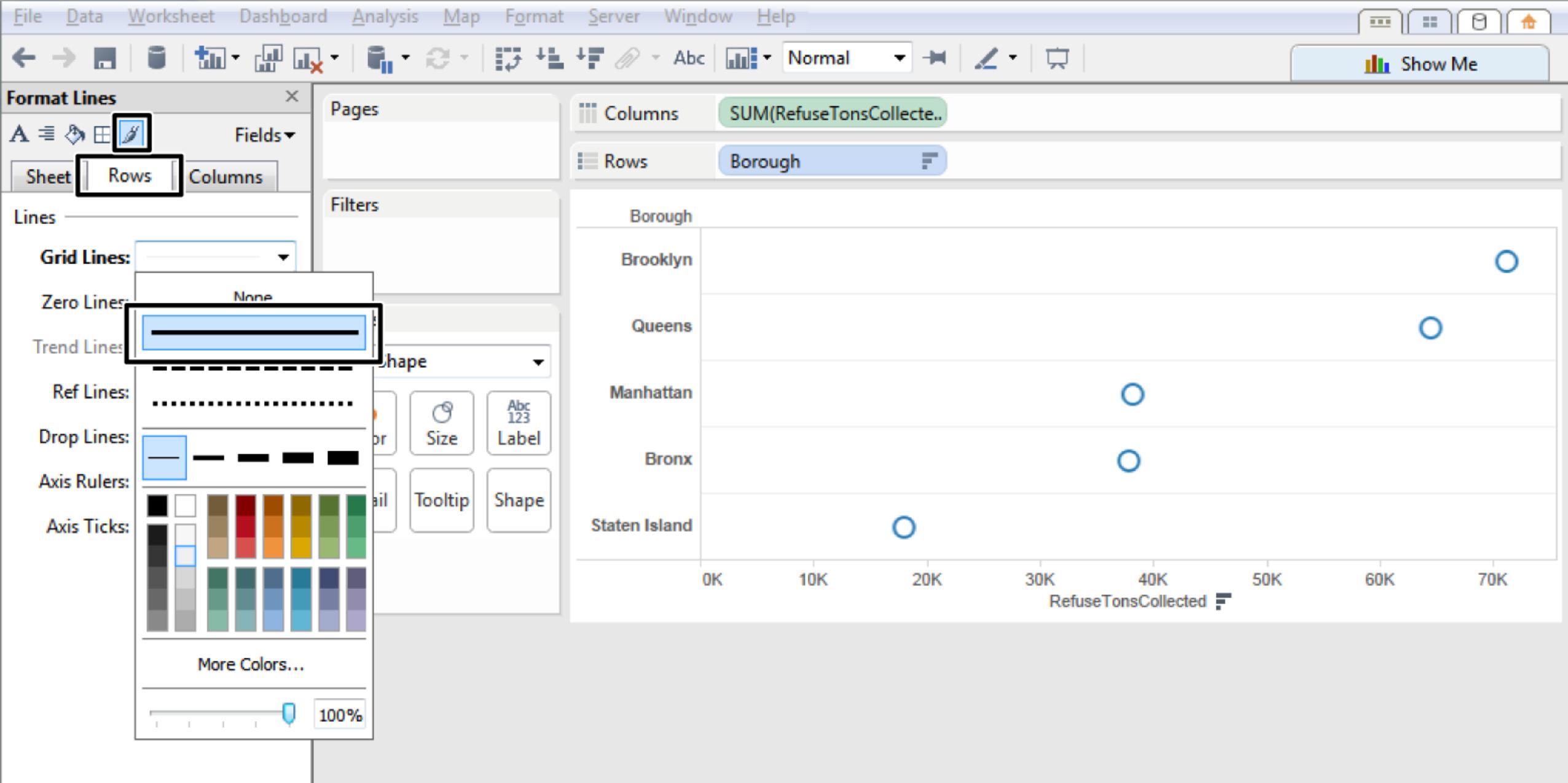


Figure 3-10. Adding grid lines to form the dot chart



# How much and How many

1

Communicating how much

2

Communicating how many

3

Ratios and Rates



# Communicating how many

How many involves countable nouns like people or distinct events:

- How many customers do we have?
- How many goals did players score this season?

- Communicating “how many” is very common and useful for data at the most granular possible level.
- For example, perhaps you want to show how many sales orders involved customers who are retired, or maybe you’re interested in presenting how many arrests were DUIs



# Communicating how many

- Let's consider an example of the same data structured two ways.
- Figure 3~12 shows the raw data for rat sightings as captured by the 311-call center in New York on the lefthand side, and a pivot table of this data set by borough on the righthand side.

A	B	C	D	E	F	G	H	I
Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip
1	25701742	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	100
2	25702233	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	100
3	25700276	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	114
5	25698779	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	100
6	25698777	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	100
7	25698869	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	114
8	25698143	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Commercial Building	112
9	25698142	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Other (Explain Below)	1120
10	25700830	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	1120
11	25701747	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Construction Site	104
12	25699463	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Mixed Use Building	1127
13	25700831	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Other (Explain Below)	1045
14	25698871	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Other (Explain Below)	112
15	25697444	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	Other (Explain Below)	1120
16	25700277	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Mixed Use Building	100
17	25698187	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Mixed Use Building	113
18	25698782	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Mixed Use Building	1120
19	25701745	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	104
20	25697428	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	112
21	25699457	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	1125
22	25701744	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	103
23	25700965	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	1031
24	25697443	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	1121
25	25698780	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	100
26	25698870	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Apt. Building	10002
27	25701517	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	1120
28	25699459	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	3+ Family Mixed Use Building	112
29	25704516	6/8/2013	DOHMH	Department of Health and Mental Hygiene	Rodent	Rat Sighting	1-2 Family Dwelling	1120

Row Labels	Count of Unique Key
BROOKLYN	11887
MANHATTAN	9328
BRONX	7188
QUEENS	5286
STATEN ISLAND	1965
Unspecified	1

Figure 3-12. Data at the incident level and aggregate level



# Communicating how many

## Counting Dimensions:

- To see how it works, connect to the rat sighting data set, and drag the Number of Records field onto the Text shelf within the Marks card.

- You will see that there are 35,655 total records in the rat sightings data set.
- If you then drag Borough onto the Rows shelf, you will reproduce the pivot table shown in Figure 3~13.

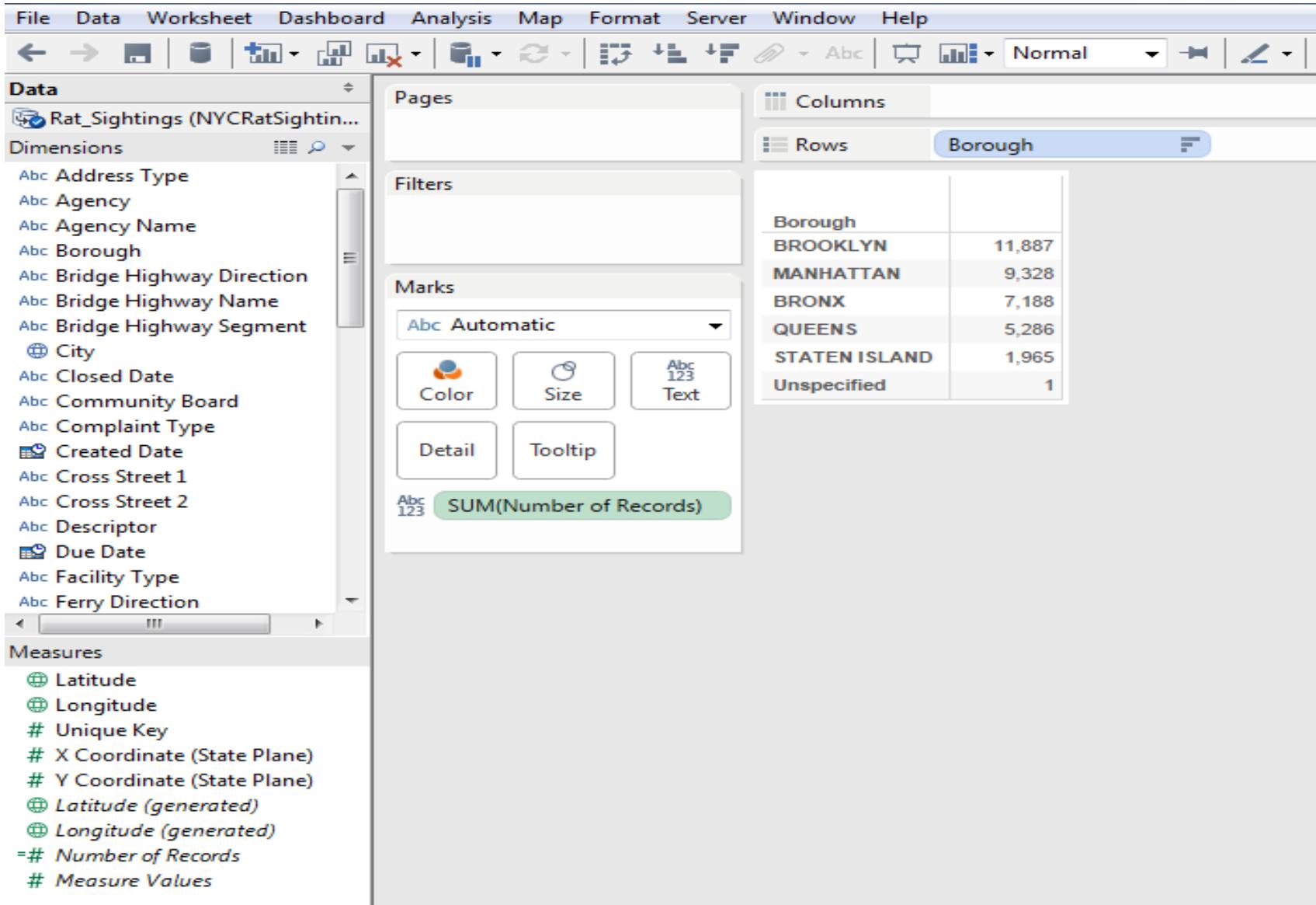


Figure 3-13. Using Number of Records to show counts



# Communicating how many

- To understand how Number of Records behaves, imagine a field in the data set of 1s.
- Summing this field for Brooklyn, for example, shows that there are 11,887 times that “Brooklyn” appears in the Borough field.
- For example, if you replace Borough in the Rows shelf with City, you will see the table shown in Figure 3-14.
- Notice anything wrong with the highlighted city name?

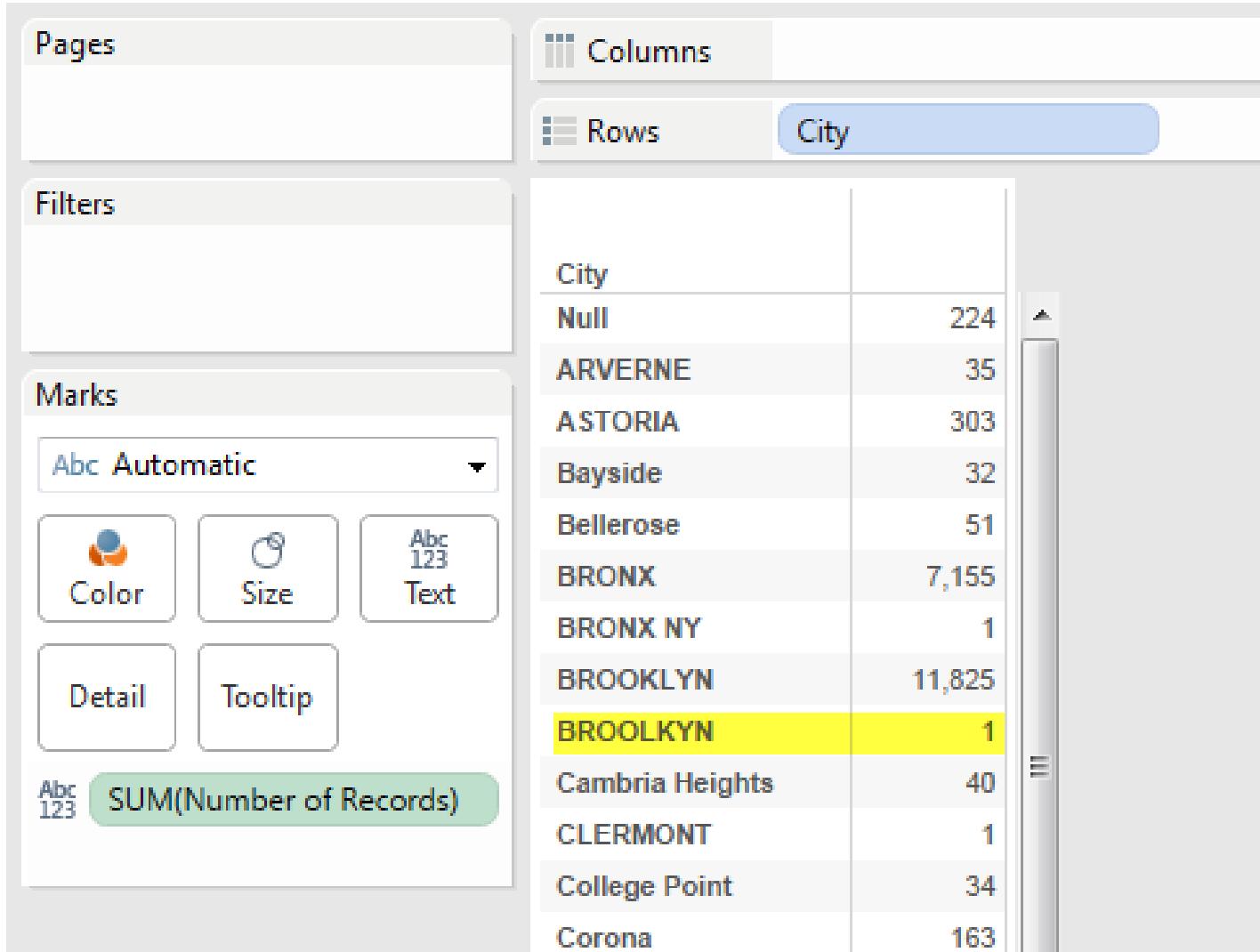


Figure 3-14. Using Number of Records to check for errors

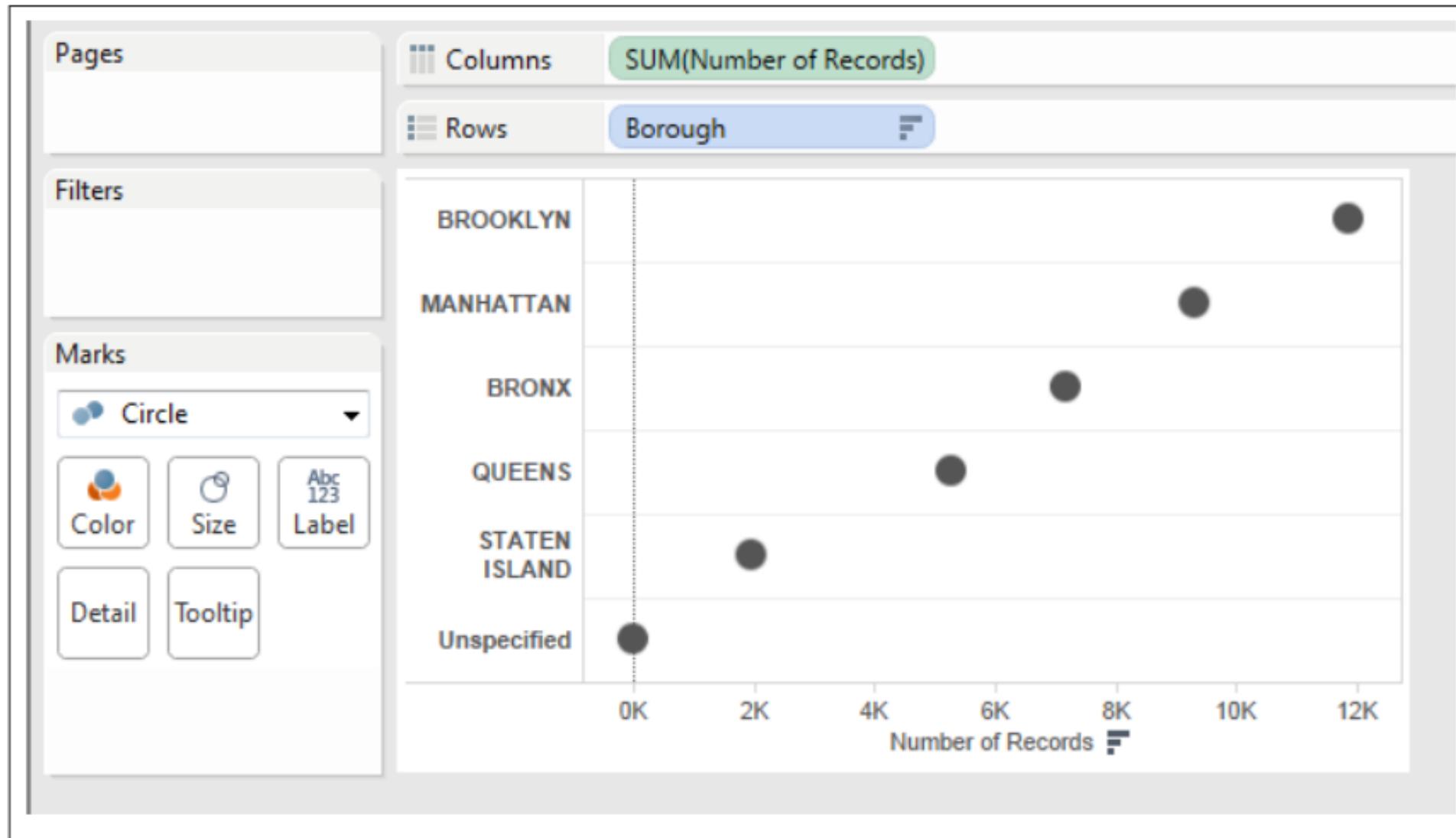


Figure 3-15. Dot chart of count of rat sightings by borough

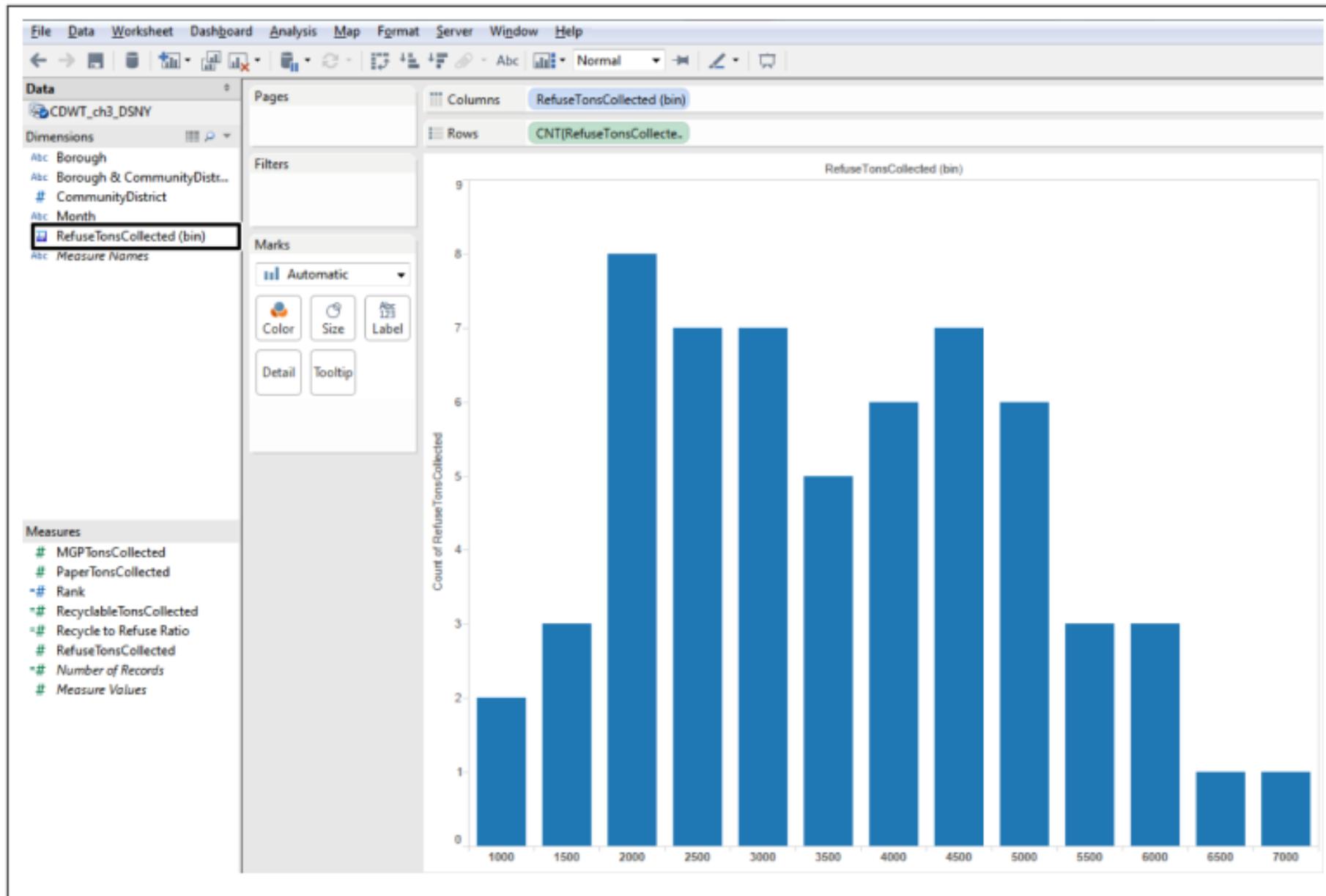


Figure 3-16. The histogram

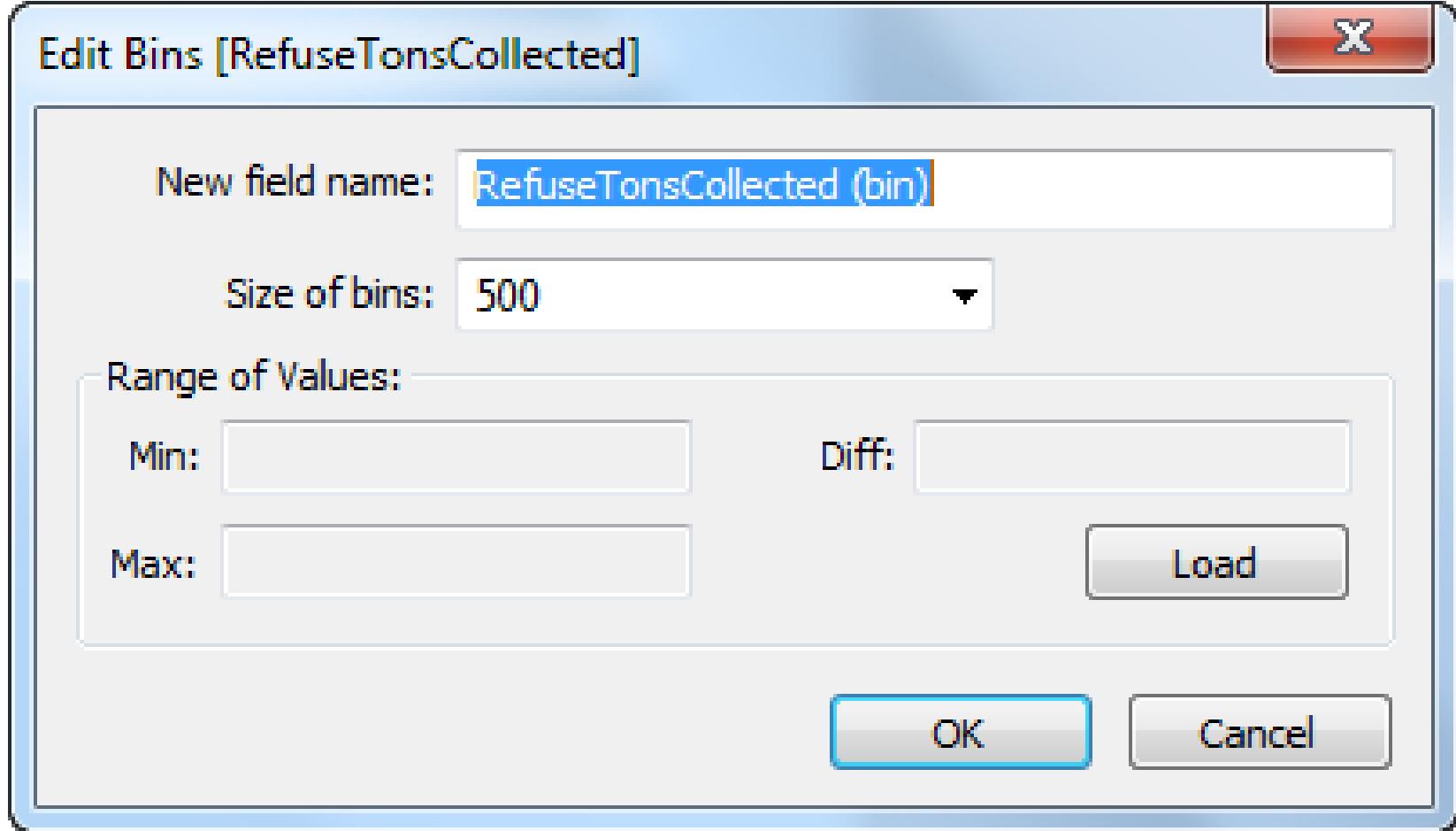


# Communicating how many

- Notice that Tableau has automatically created a new field in the Dimensions area called RefuseTonsCollected (bin).
- This data field type is used to determine the discrete intervals into which the individual records are bucketed.

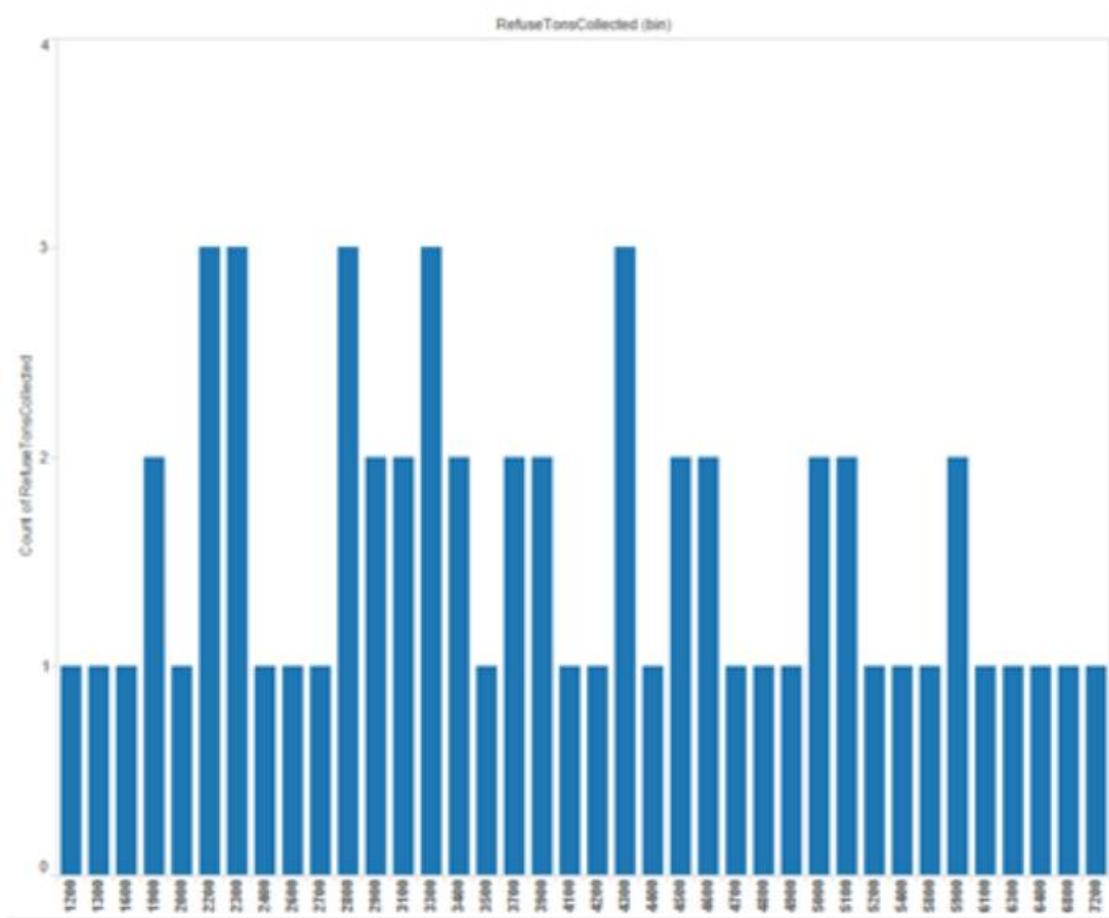
How can we read this chart?

- The first bar of the chart shows that there were two records that produced between 1,000 and 1,500 tons of garbage.
- The second bar, with a height of three, is for community districts producing between 1,500 and 2,000 tons of garbage, and so on.
- The bin size is therefore 500 tons.

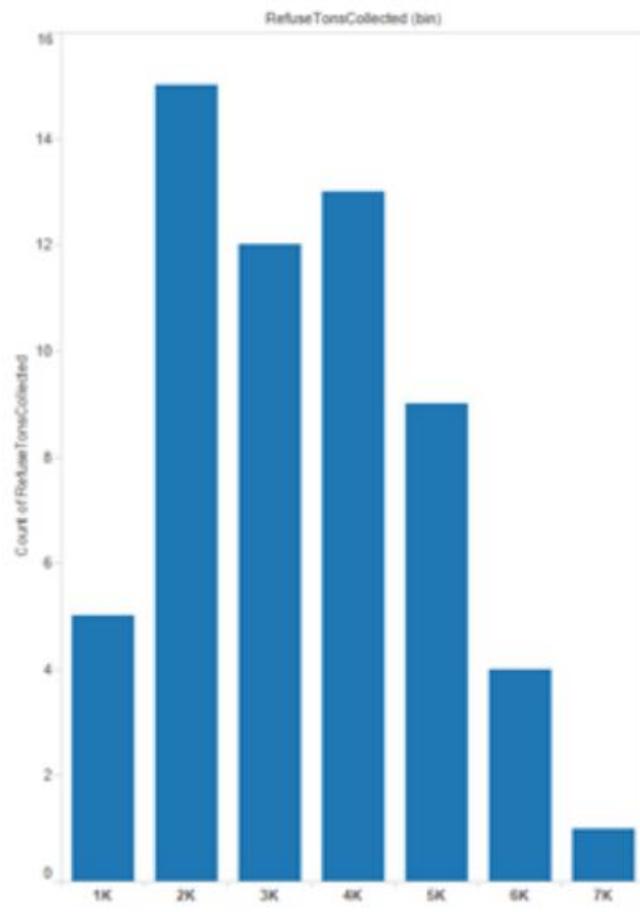


*Figure 3-17. The Edit dialog box for the histogram bin field*

**Size of bin = 100**



**Size of bin = 1,000**

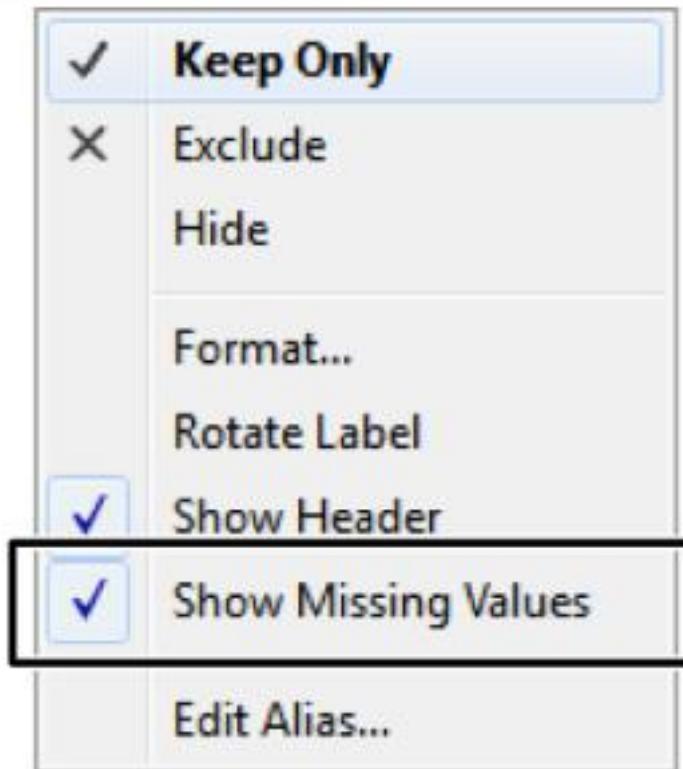


*Figure 3-18. Histograms with different bin sizes*

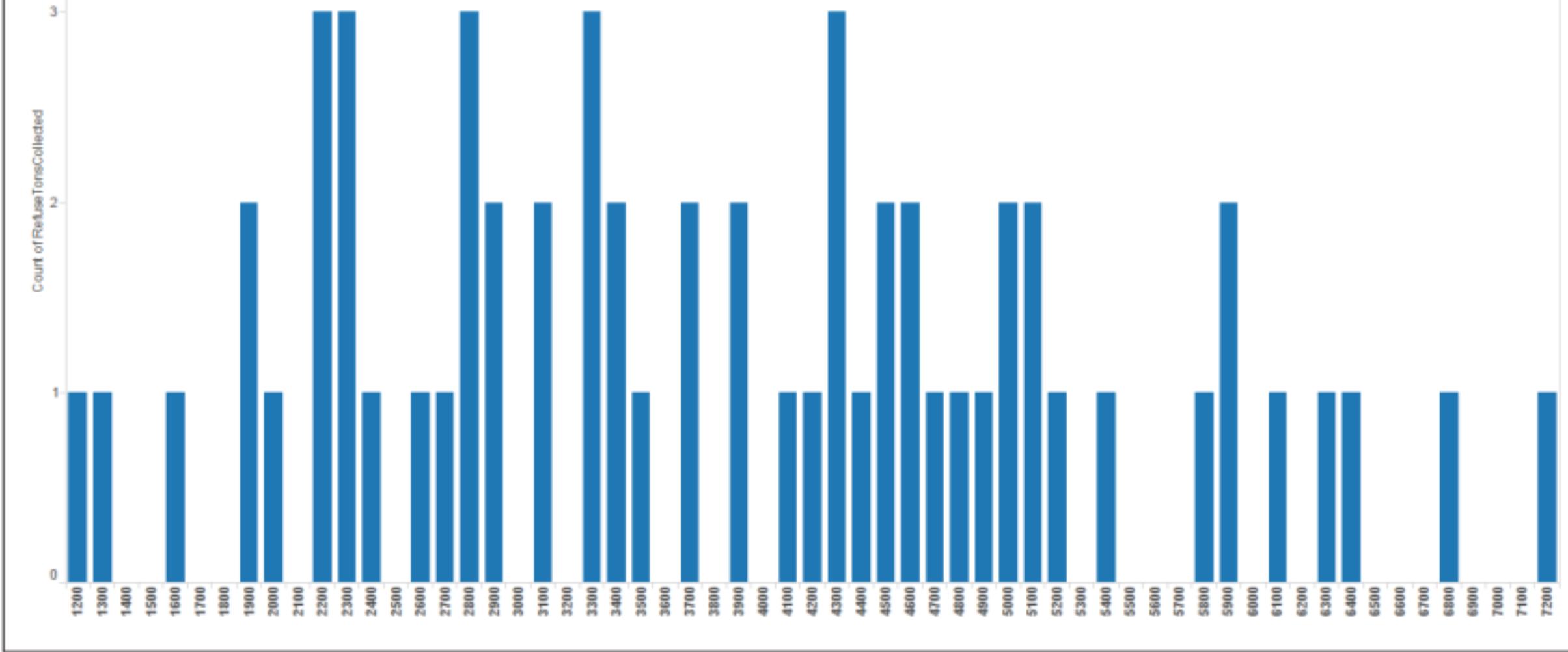


# Communicating how many

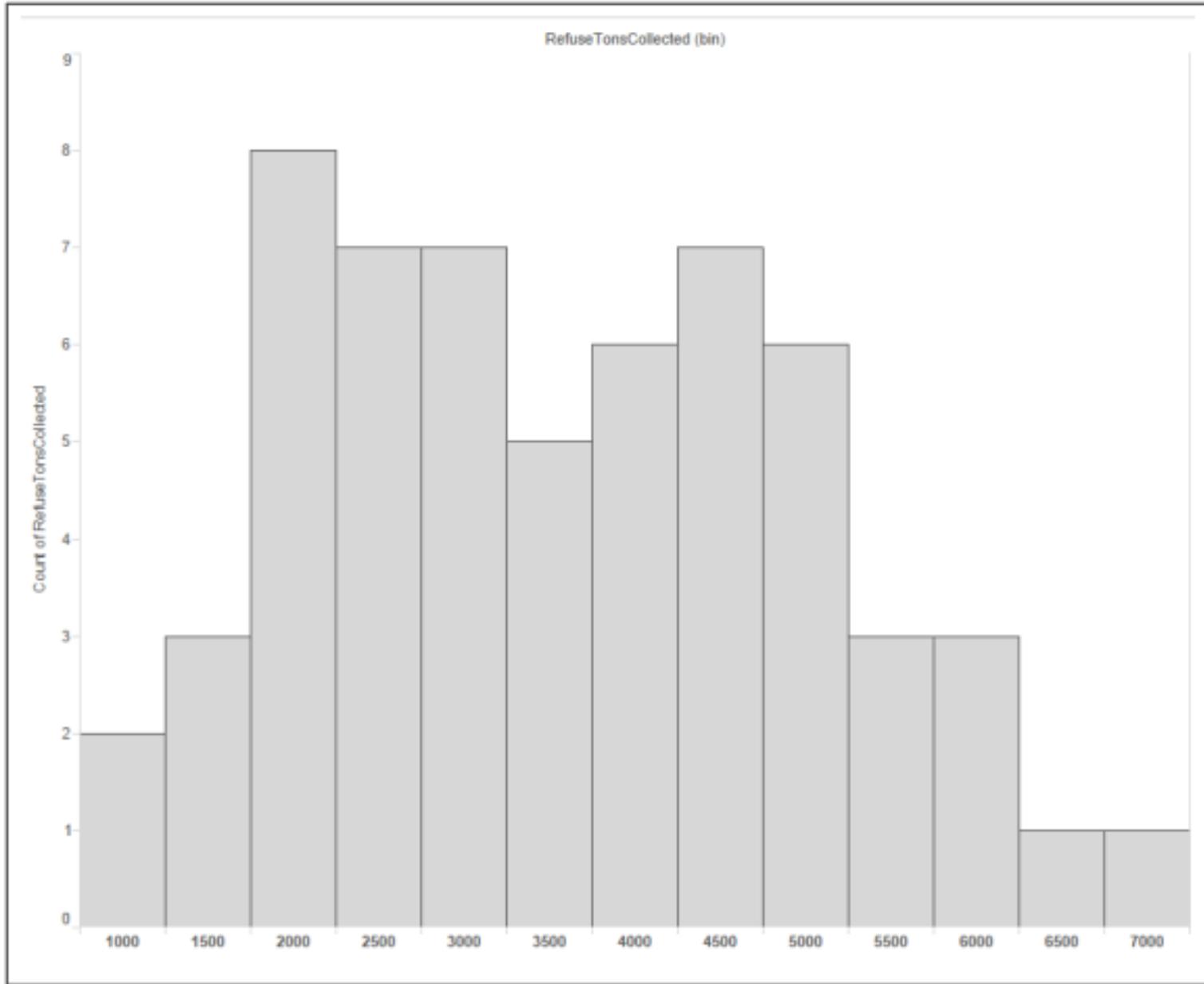
- There's one thing to watch out for with histograms made in Tableau.
- If you look closely at the histogram with bins sized at 100 tons, you'll notice that some intervals are missing.
- The first bar is for 1,200–1,300 tons and the second is for 1,300–1,400 tons. You'd expect the third bar to be for 1,400–1,500 tons, but it isn't. It's for 1,600–1,700 tons. Why is this?



*Figure 3-19. The Show Missing Values box for histograms*



*Figure 3-20. Histograms should be shown with missing values*



*Figure 3-21. A histogram with improved formatting*



# How much and How many

1

Communicating how much

2

Communicating how many

3

Ratios and Rates



# Ratios and Rates

- A **ratio** is a comparison of two terms expressed as a quotient.
- For example, Manhattan produced 0.264 tons of recycle for every ton of refuse. Ratios can be expressed as “x to y,” “x:y,” “x/y,” or as a decimal.

- A **rate** is a ratio in which the two terms have different units.
- For example, the population density of Brooklyn is 36,136 residents per square mile. Rates are often predictive because time can be used as the denominator (crime rates, population growth rates, etc.).



# Ratios and Rates

- A **proportion** is a ratio in which the numerator is a partial amount, and the denominator is the total amount (expressed as a number between 0 and 1).  
• For example, the proportion of the NYC population living in the Bronx is 0.169. A proportion is expressed as a number between 0 and 1.
- A **percentage** is a ratio comparing a number to 100.  
• For example, 16.9% of NYC residents live in the Bronx. A percentage is generally a number between 0 and 100, but can be larger than 100 (e.g., “sales have increased by 150% year-over-year”).

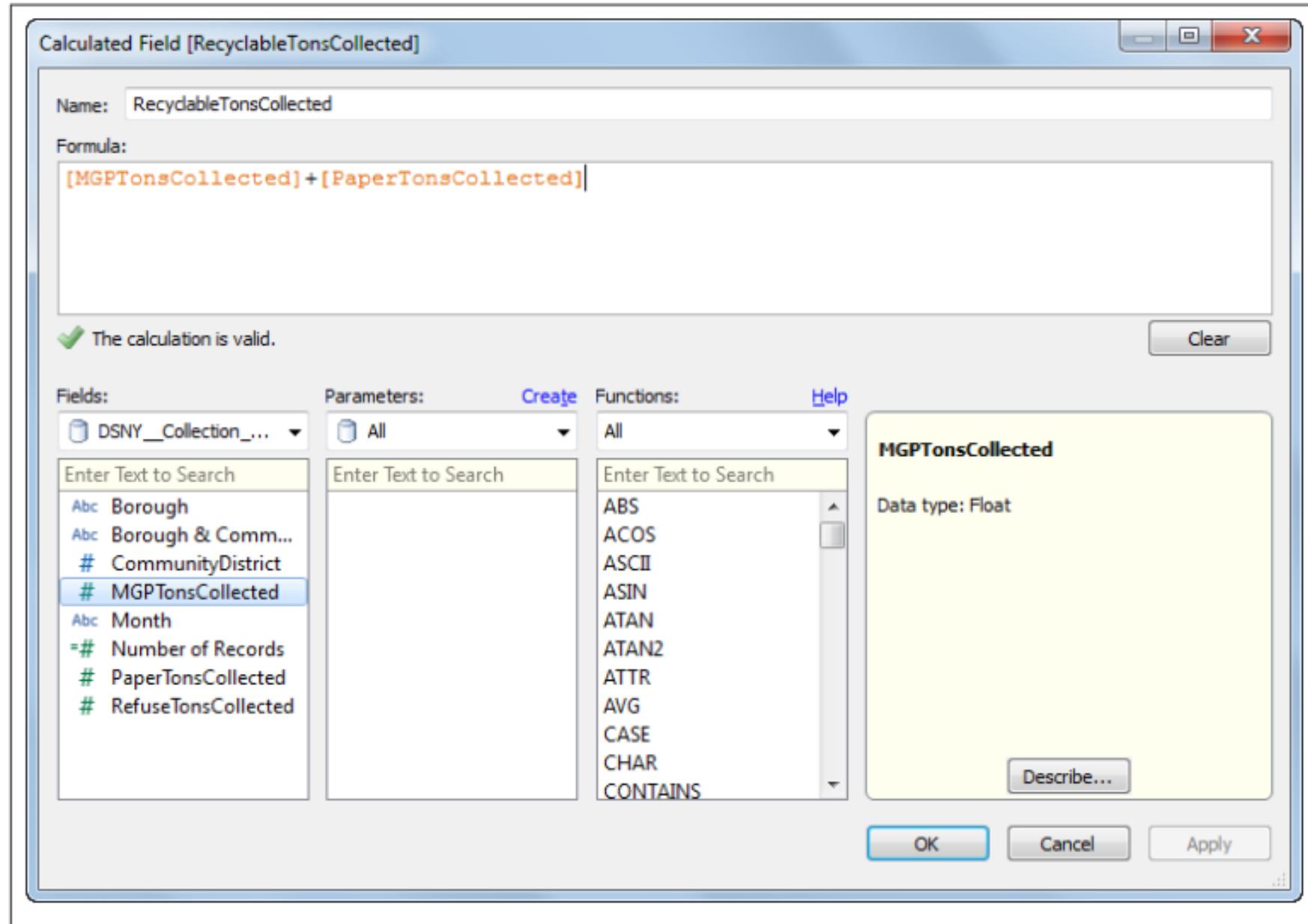


Figure 4-1. Summing two fields with a Calculated Field

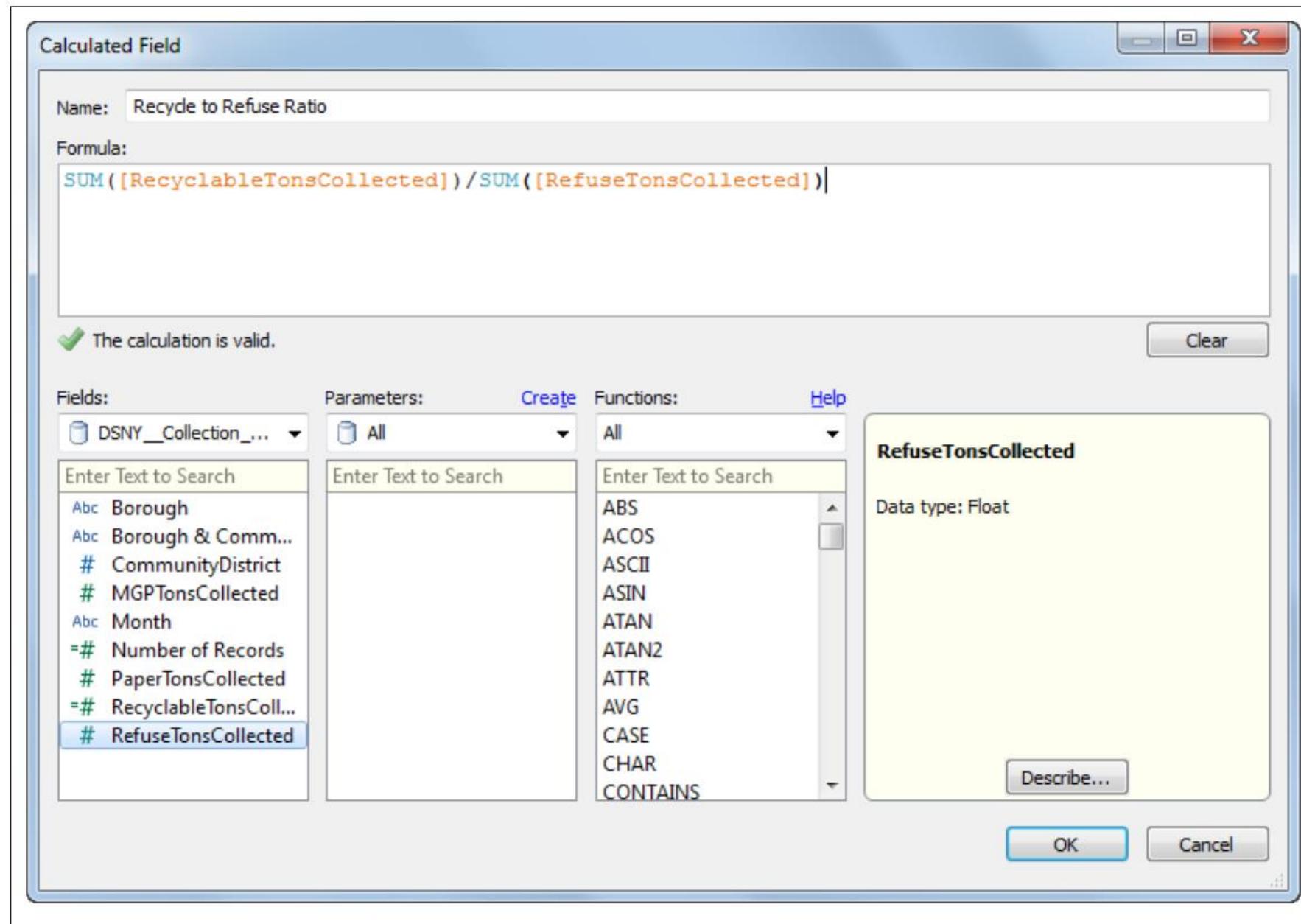


Figure 4-2. Creating a ratio with a Calculated Field



Figure 4-4. Ratios by community district and borough

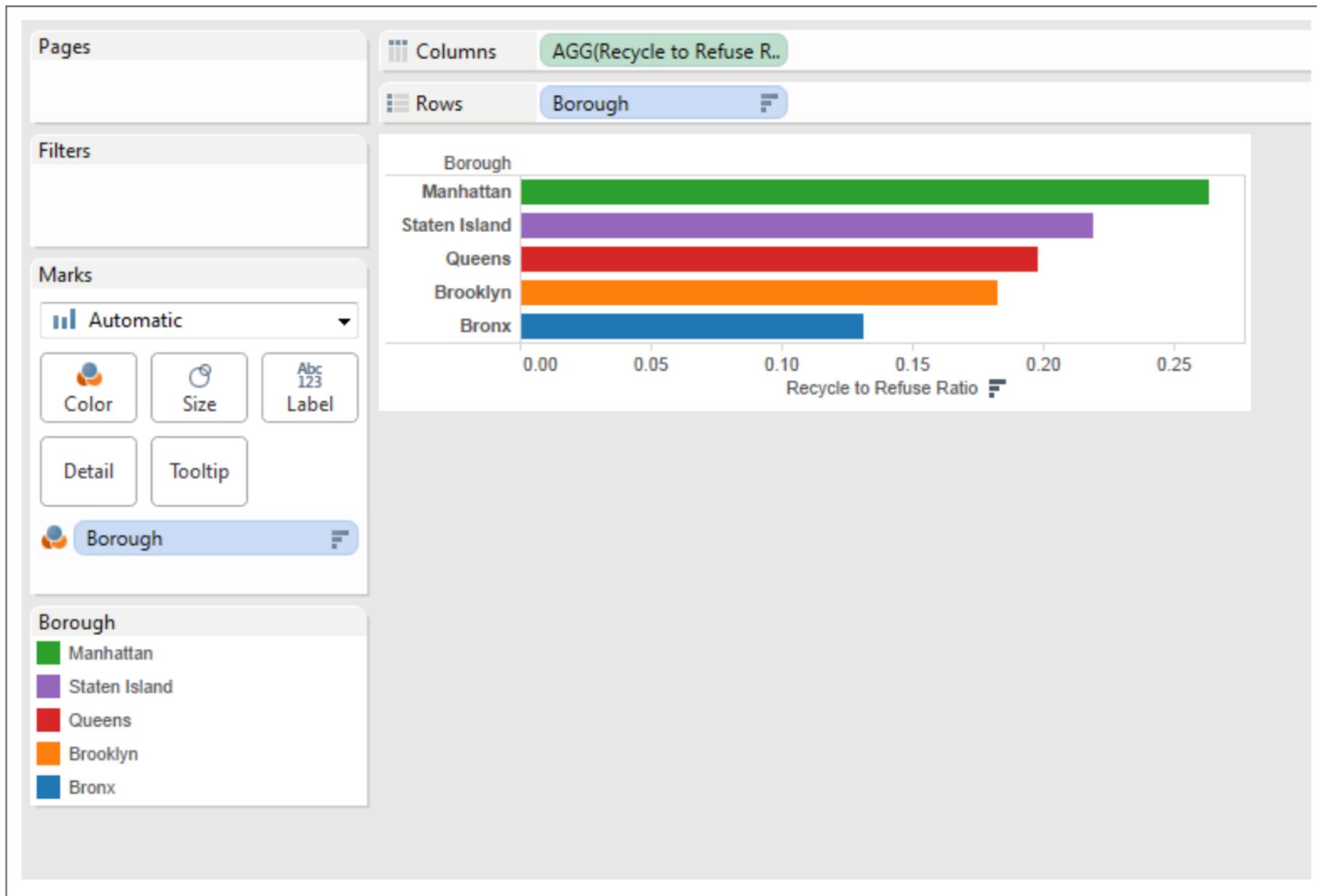


Figure 4-3. Visualizing the ratio of recycle to refuse for each borough

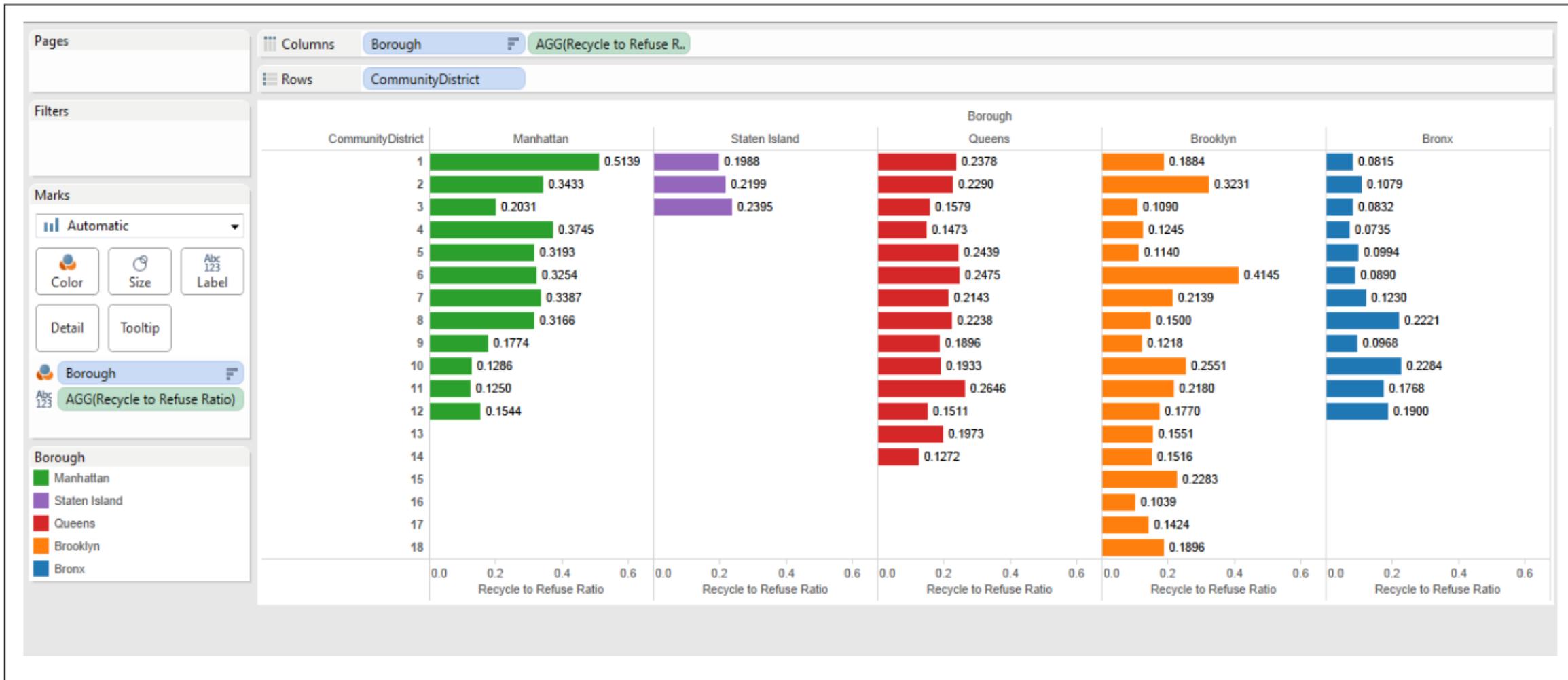


Figure 4-5. Adding labels to the bar chart grid

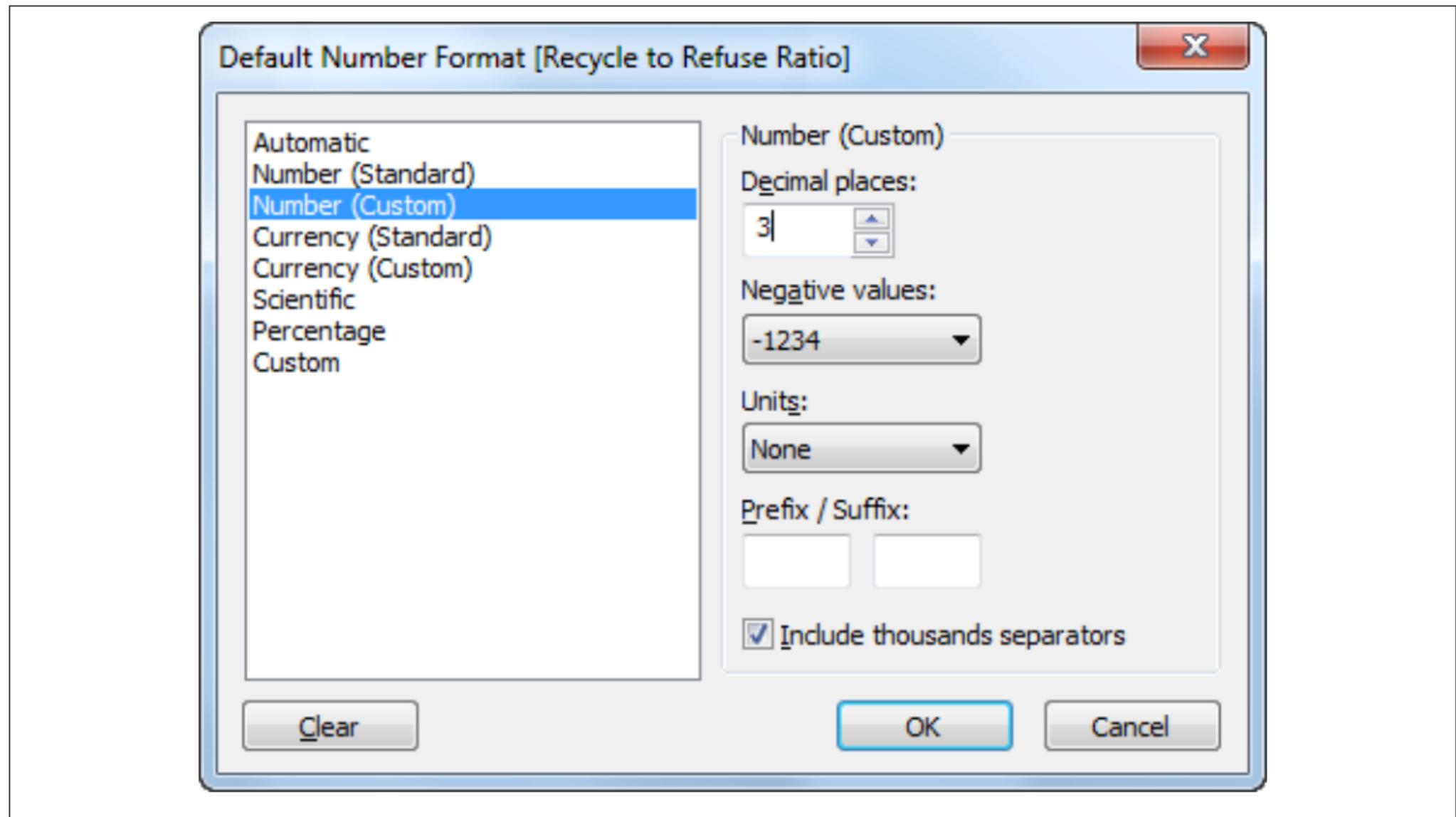


Figure 4-6. Specifying the number of decimal places in the Measures panel

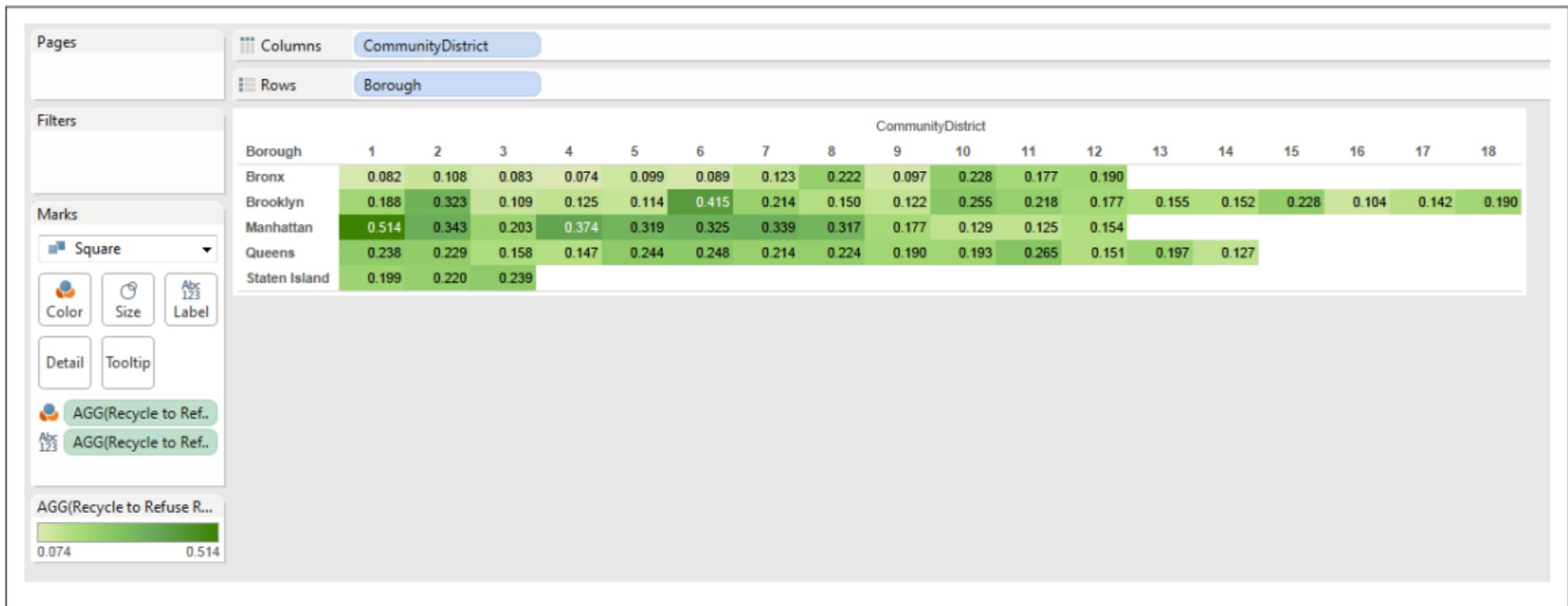


Figure 4-7. The default highlight table for Recycle to Refuse Ratio

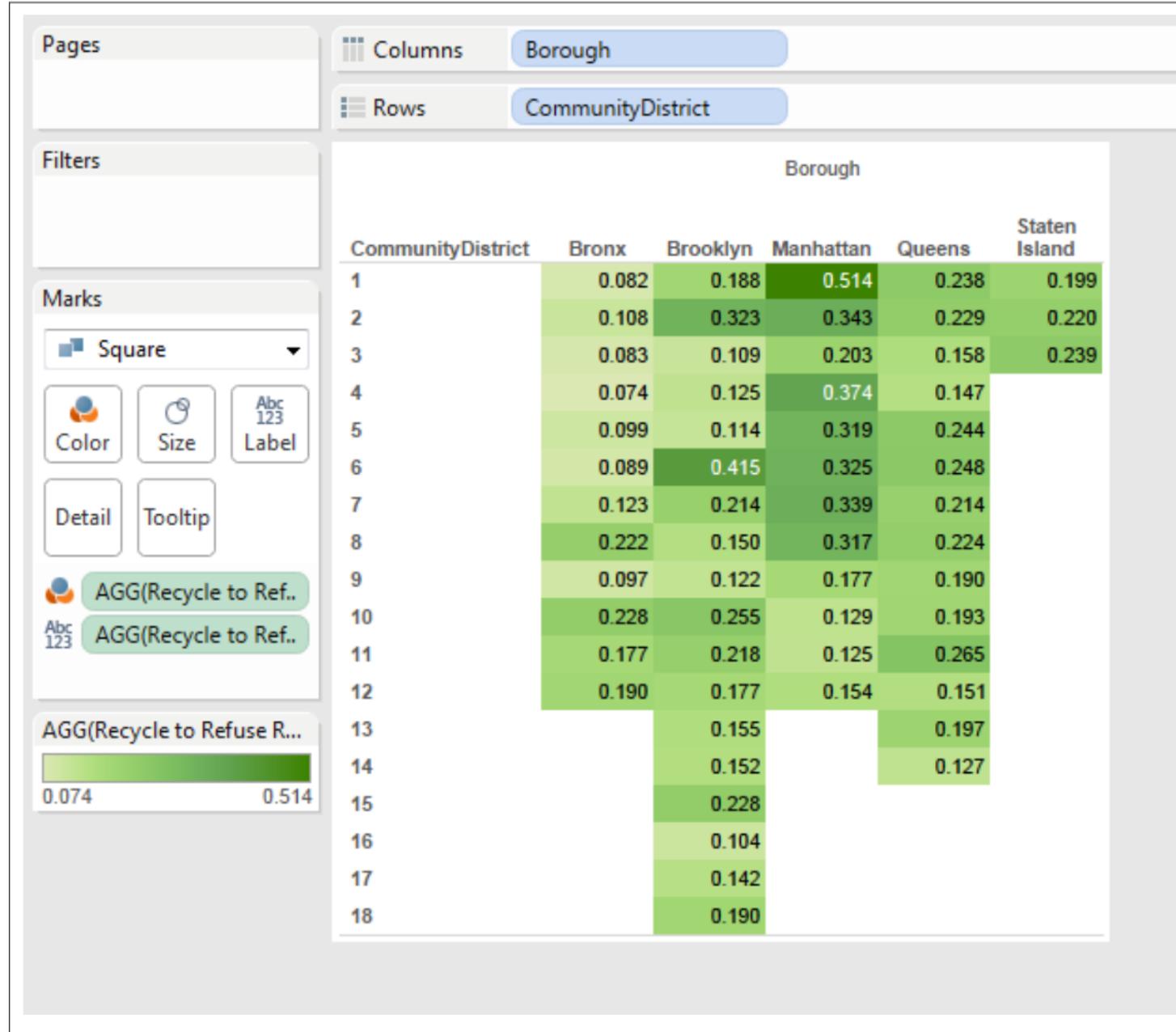


Figure 4-8. The highlight table with Rows and Columns swapped

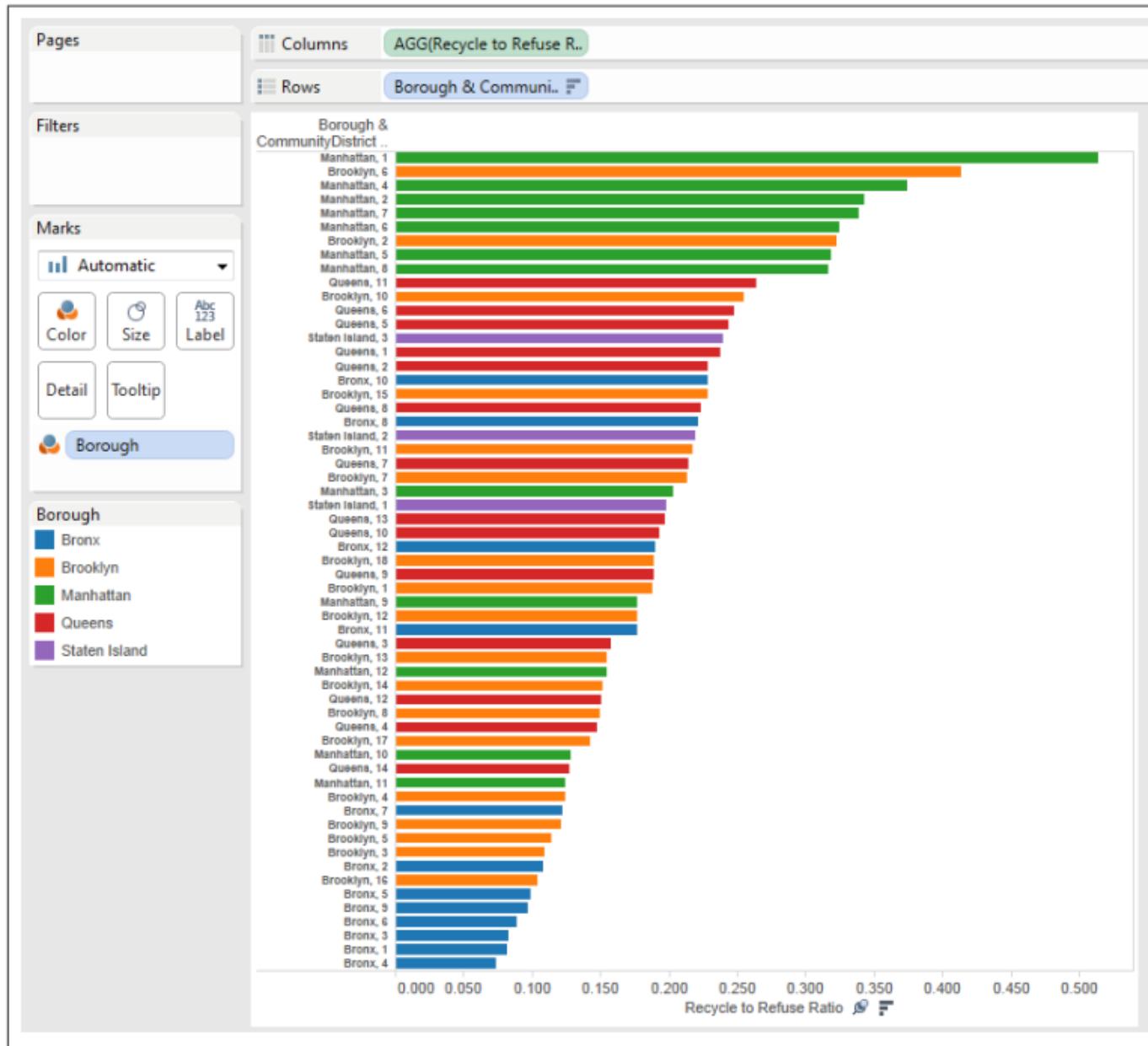


Figure 4-9. Bar chart of community districts

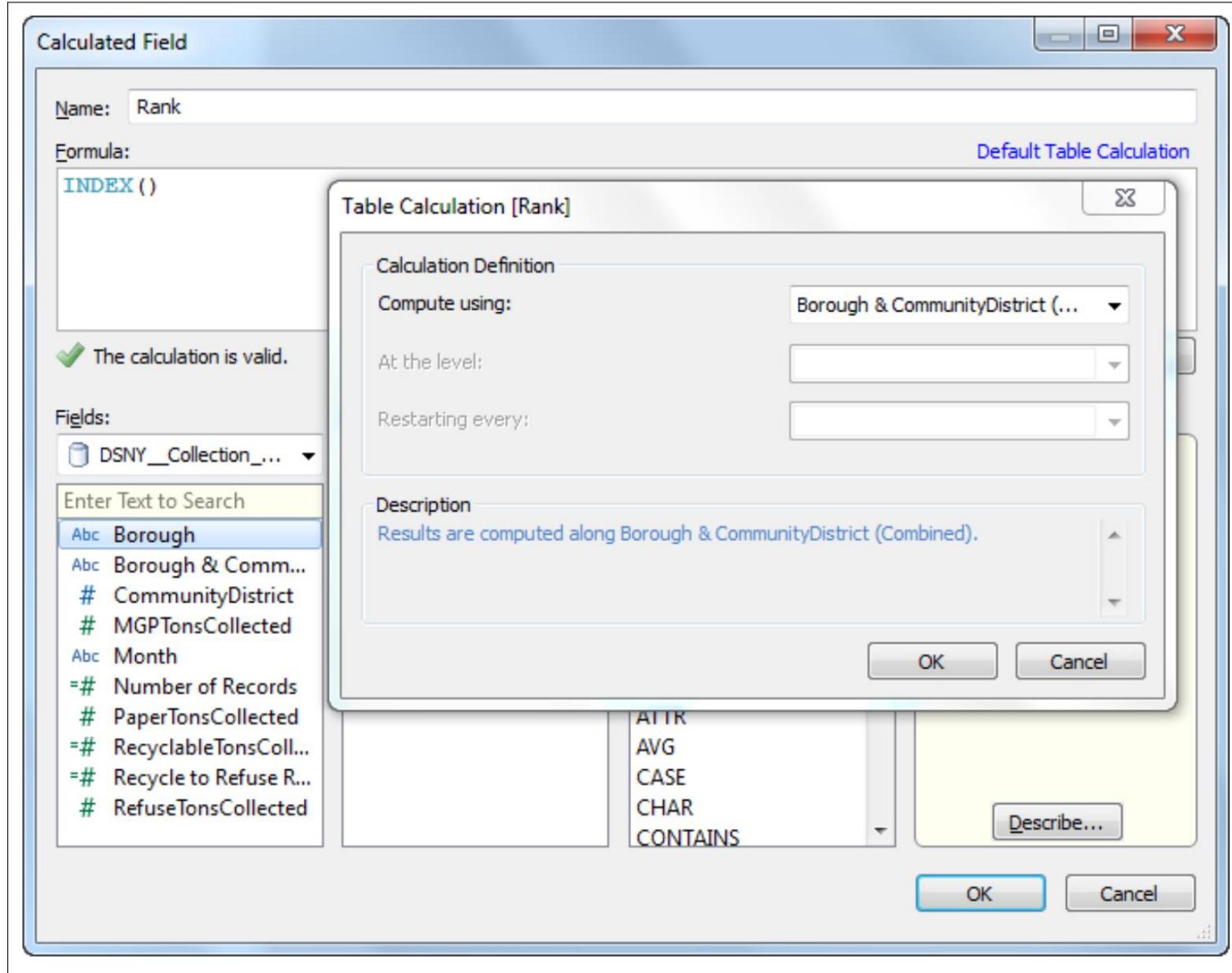


Figure 4-10. Creating a rank for each of the 59 community districts

Pages

Columns

AGG(Recycle to Refuse R...

Rows

Rank



Borough & Communi.. F

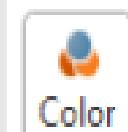
Filters

Borough &  
Rank CommunityDistrict..

- |    |                  |
|----|------------------|
| 1  | Manhattan, 1     |
| 2  | Brooklyn, 6      |
| 3  | Manhattan, 4     |
| 4  | Manhattan, 2     |
| 5  | Manhattan, 7     |
| 6  | Manhattan, 6     |
| 7  | Brooklyn, 2      |
| 8  | Manhattan, 5     |
| 9  | Manhattan, 8     |
| 10 | Queens, 11       |
| 11 | Brooklyn, 10     |
| 12 | Queens, 6        |
| 13 | Queens, 5        |
| 14 | Staten Island, 3 |
| 15 | Queens, 1        |
| 16 | Queens, 2        |
| 17 | Bronx, 10        |

Marks

Automatic



Color



Size



Label

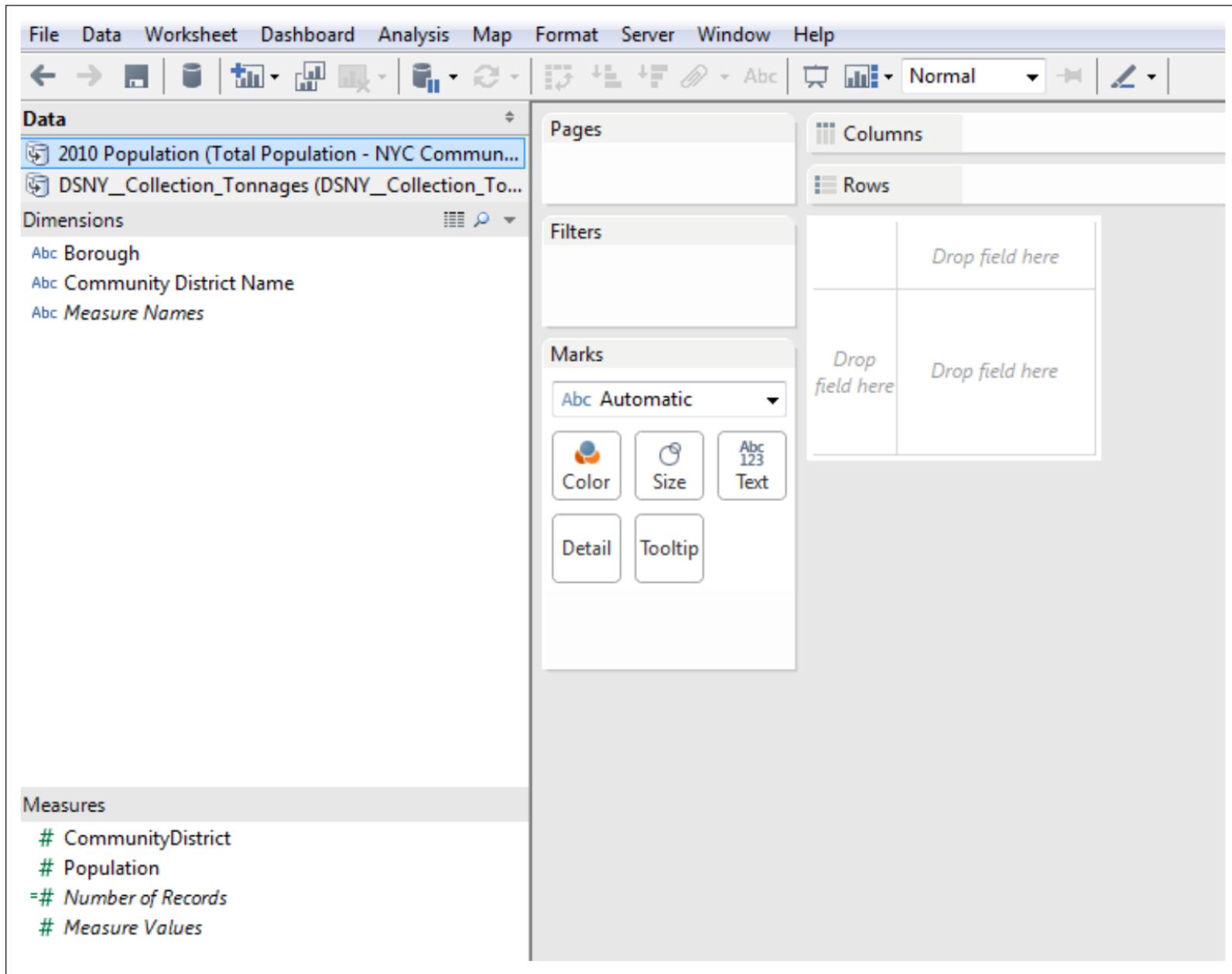


Figure 4-13. Adding a second data source to a workbook

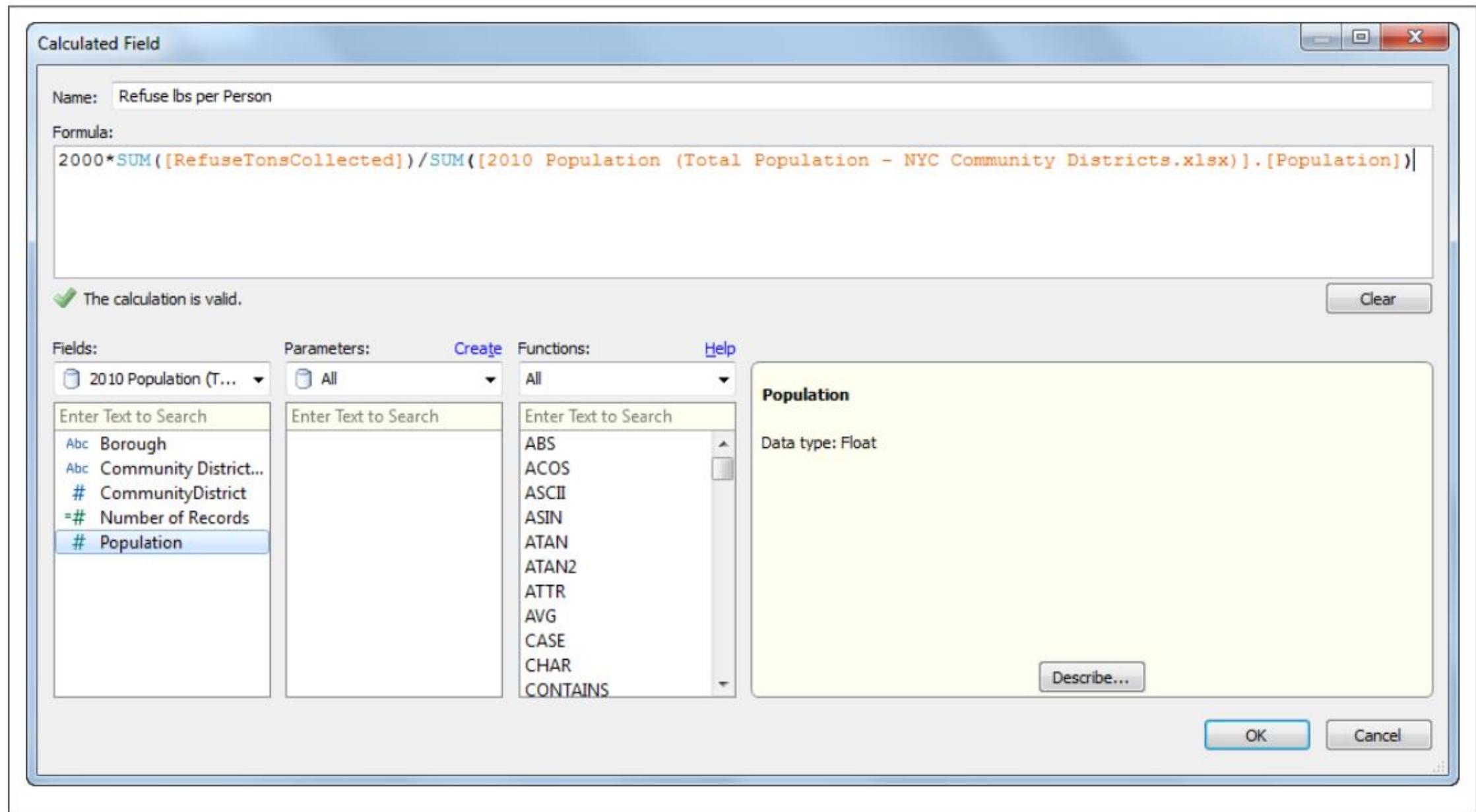
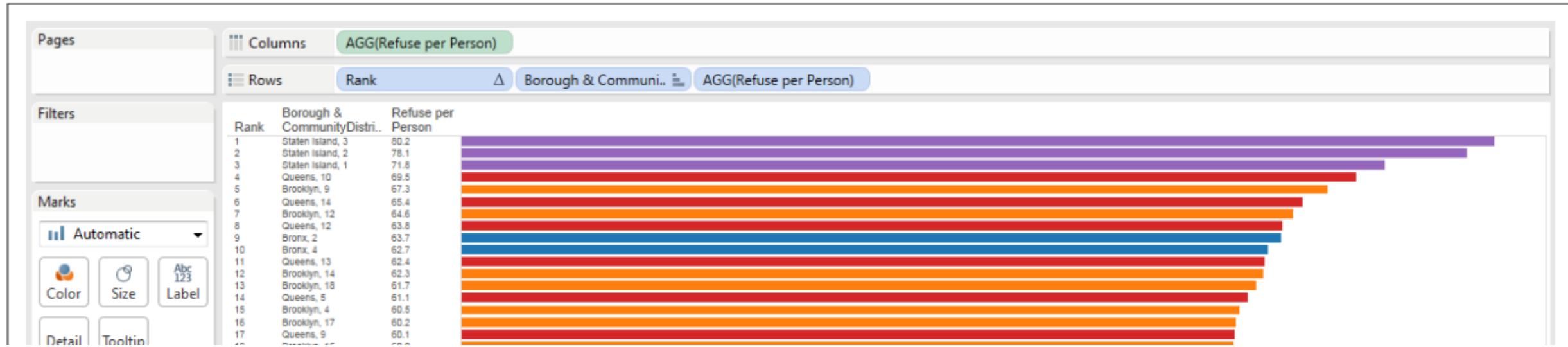


Figure 4-14. Creating a rate using fields from two data sources





# How much and How many

1

Communicating how much

2

Communicating how many

3

Ratios and Rates



# DATA VISUALIZATION



# **UNIT - IV (PART - I)**



# Proportion & Percentage

- A **proportion** is a ratio in which the numerator is a partial amount, and the denominator is the total amount (expressed as a number between 0 and 1).  
• For example, the proportion of the NYC population living in the Bronx is 0.169. A proportion is expressed as a number between 0 and 1.
- A **percentage** is a ratio comparing a number to 100.  
• For example, 16.9% of NYC residents live in the Bronx. A percentage is generally a number between 0 and 100, but can be larger than 100 (e.g., “sales have increased by 150% year-over-year”).



# A step in the process

- They are most often used to communicate three different types of comparisons:
  1. Part-to-whole
  2. Current-to-historical
  3. Actual-to-target



# Proportions and Percentages

1

Part to whole

2

Current to historical

3

Actual to target

4

Mean and Median



# Part to whole

- A part to whole analogy is defined as a comparison between a part and a whole of one thing and how it is like a part and a whole of another thing
- For example, Fin : Fish :: Wing : Bird is a part to whole analogy and works because it demonstrates that just as a fish uses a fin to move, so too does a bird use a wing to move.
- Followers of the sport of baseball have long been fascinated with proportions of outcomes (called stats).

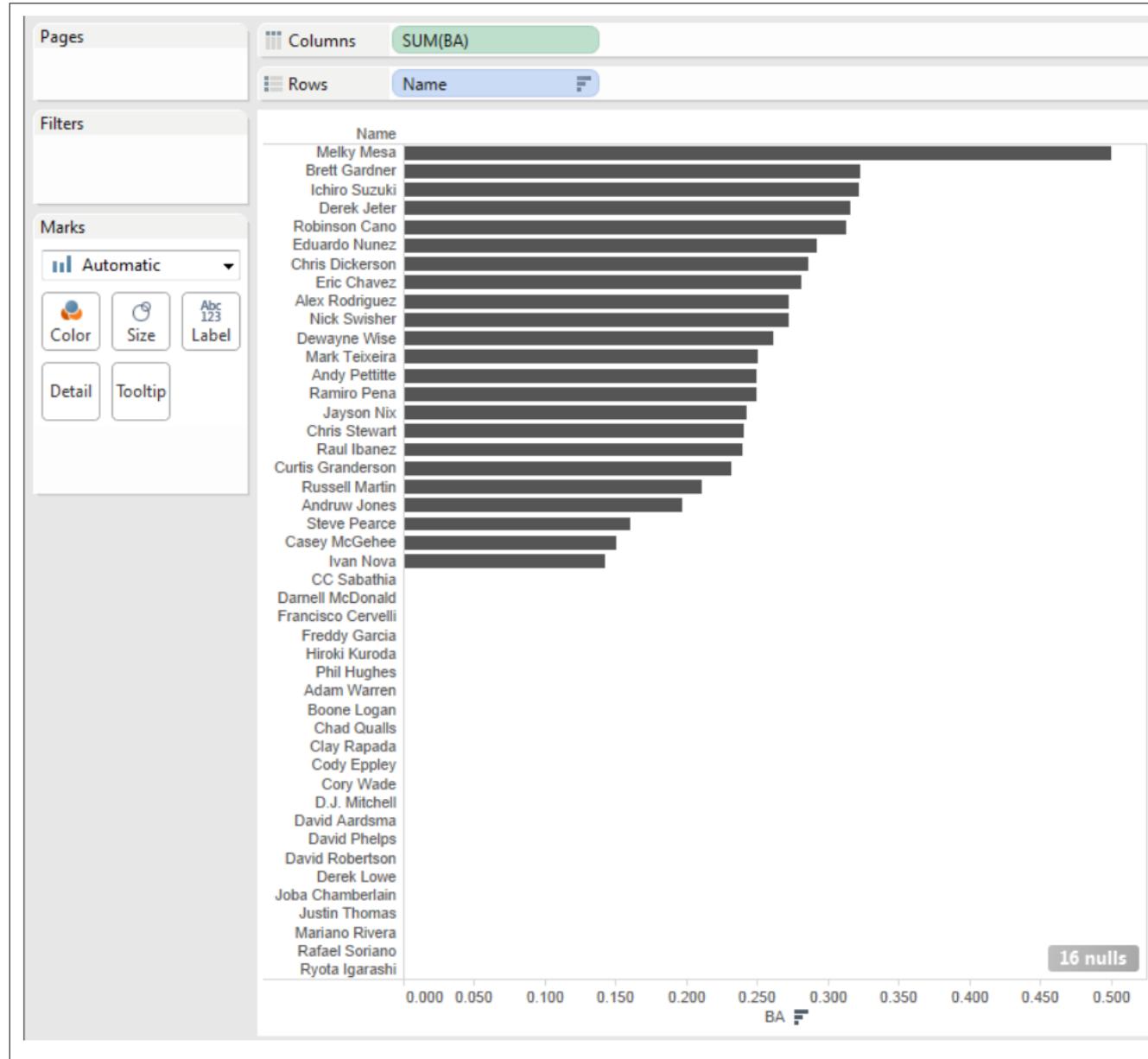


Figure 5-1. Bar chart showing BA for every Yankee with at least one at bat in 2012

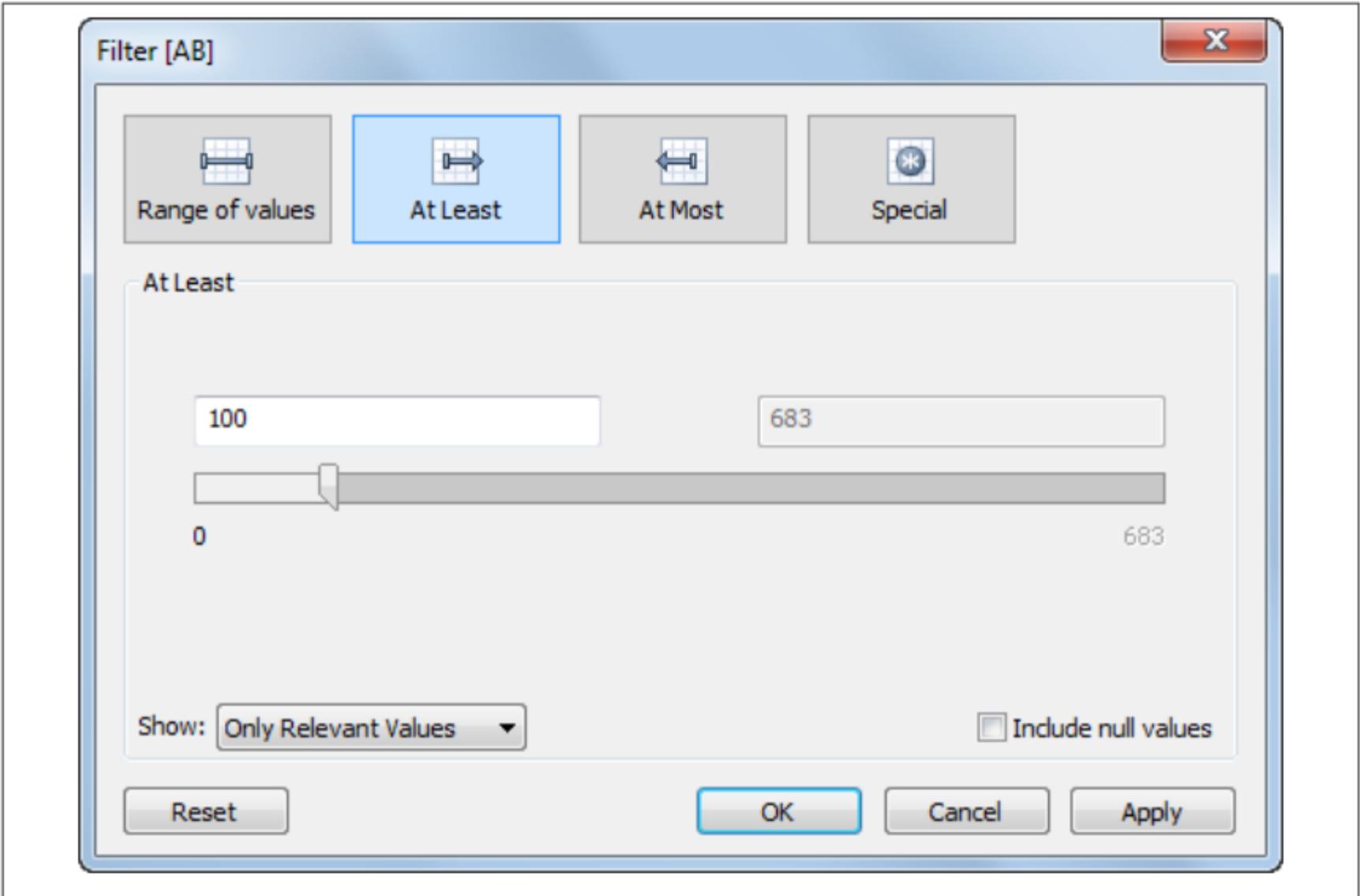


Figure 5-2. Filtering the BA bar chart to show only players with at least 100 at bats

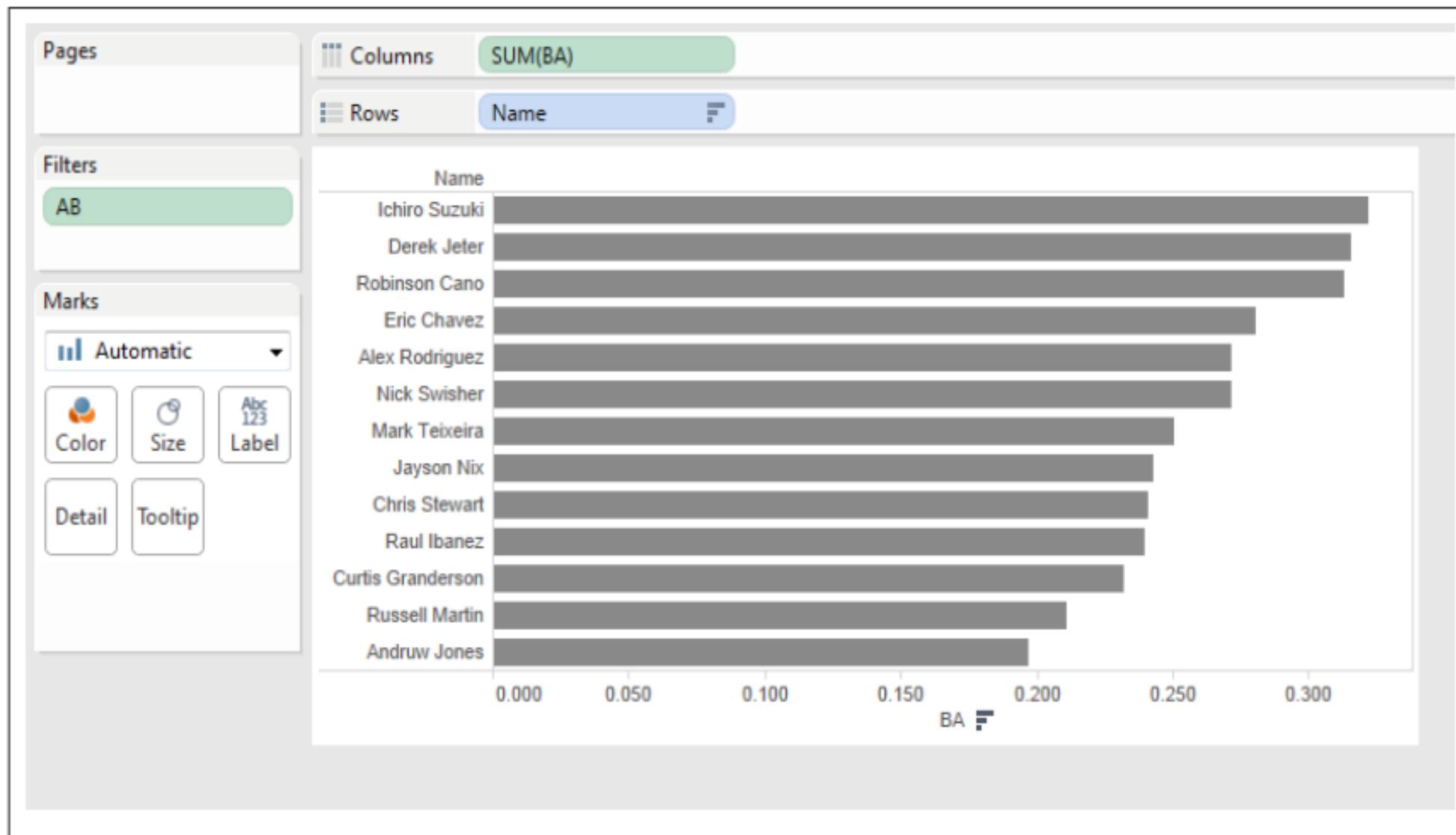


Figure 5-3. The filtered BA bar chart

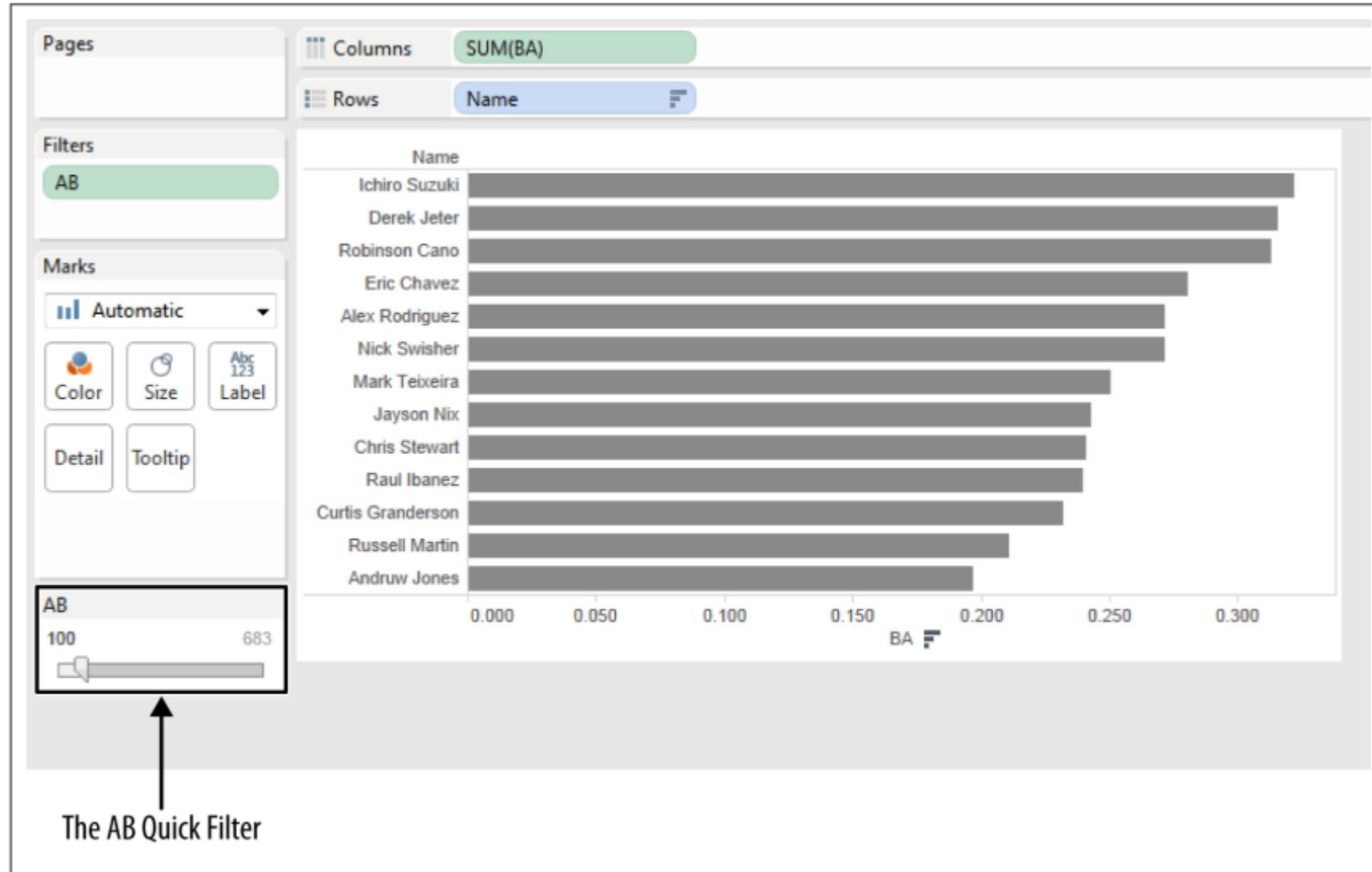


Figure 5-4. The AB Quick Filter added to the Sheet

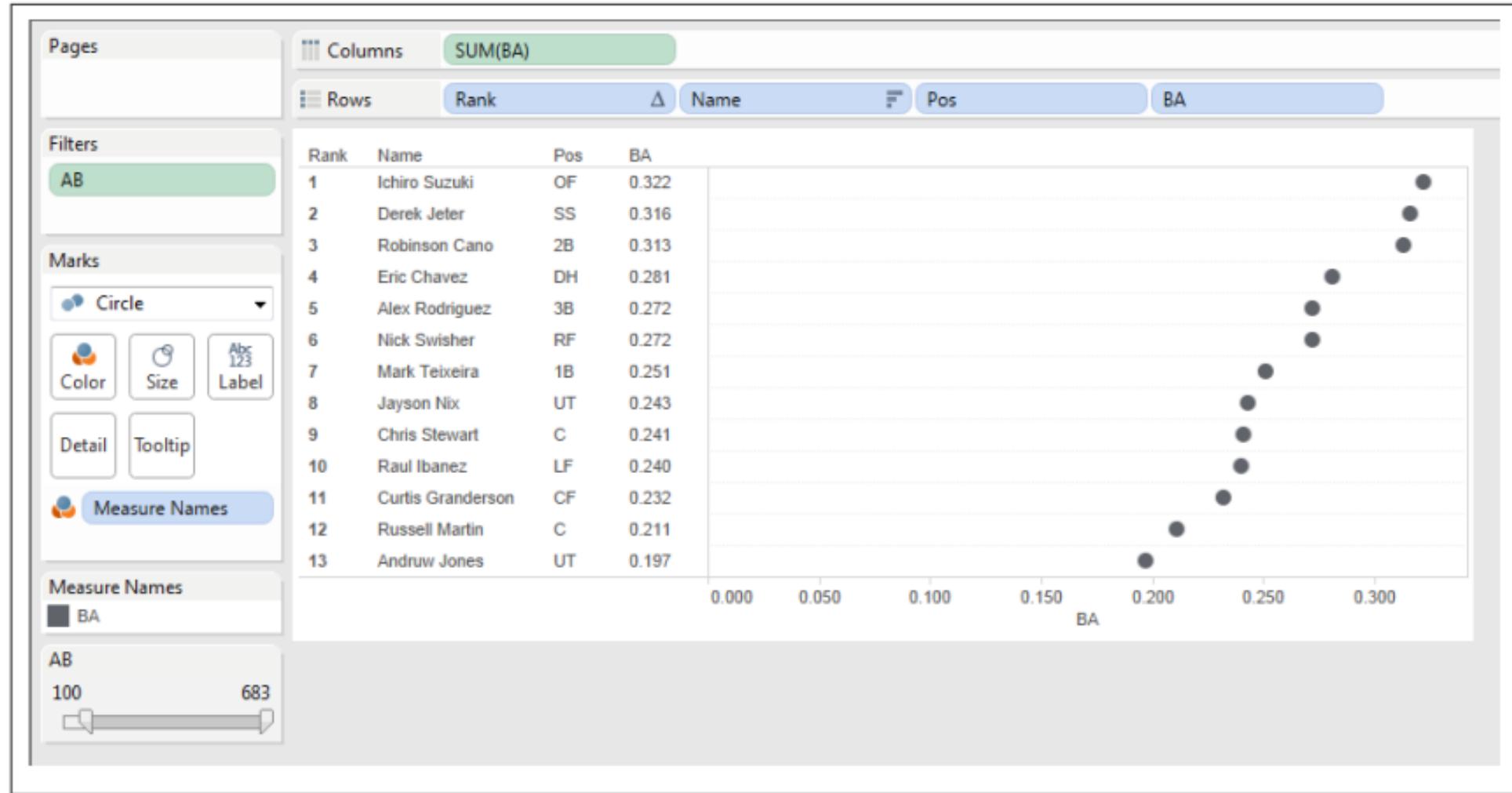
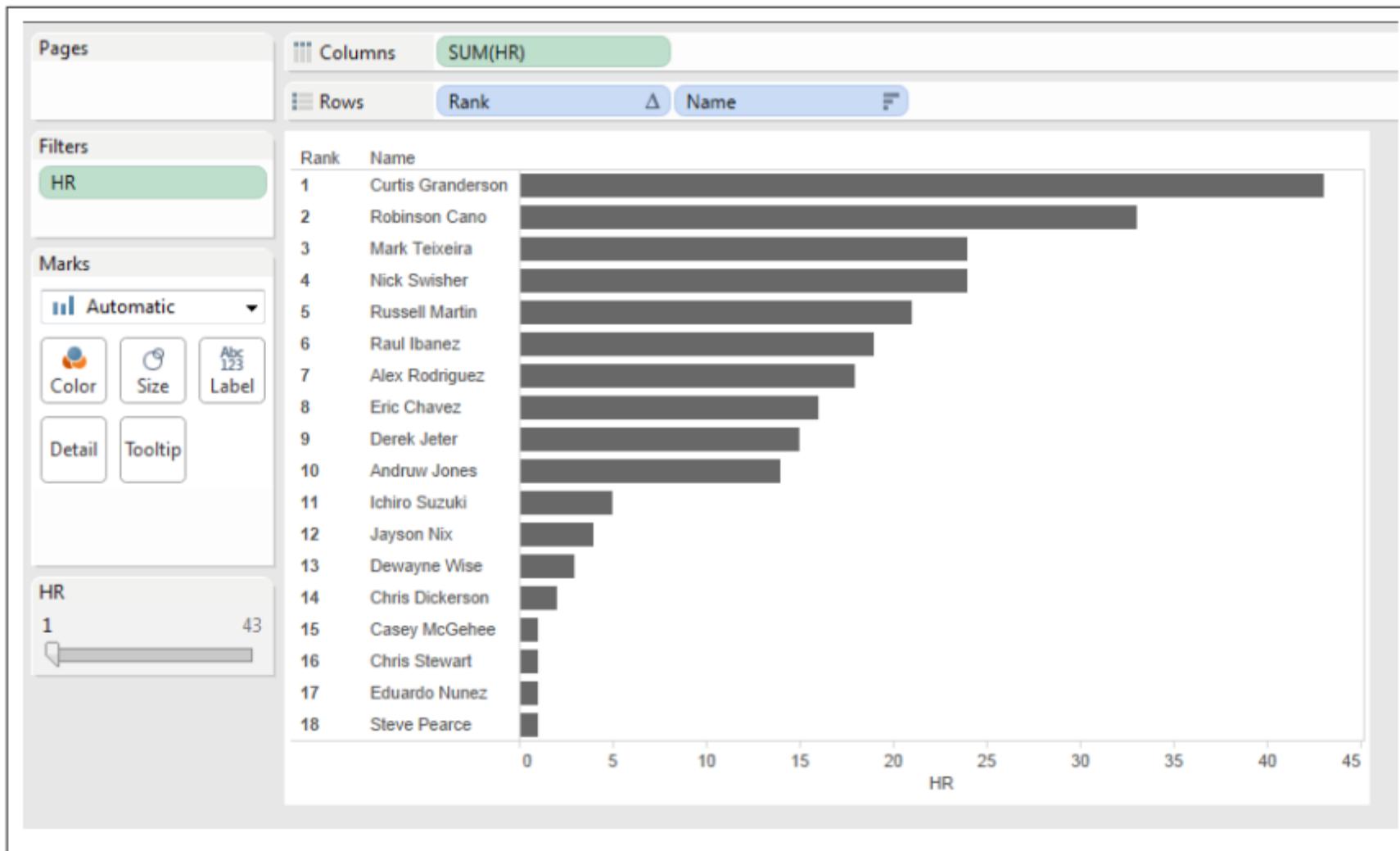


Figure 5-5. Batting average dot chart



*Figure 5-6. Bar chart showing Yankees in 2012 with at least one home run*

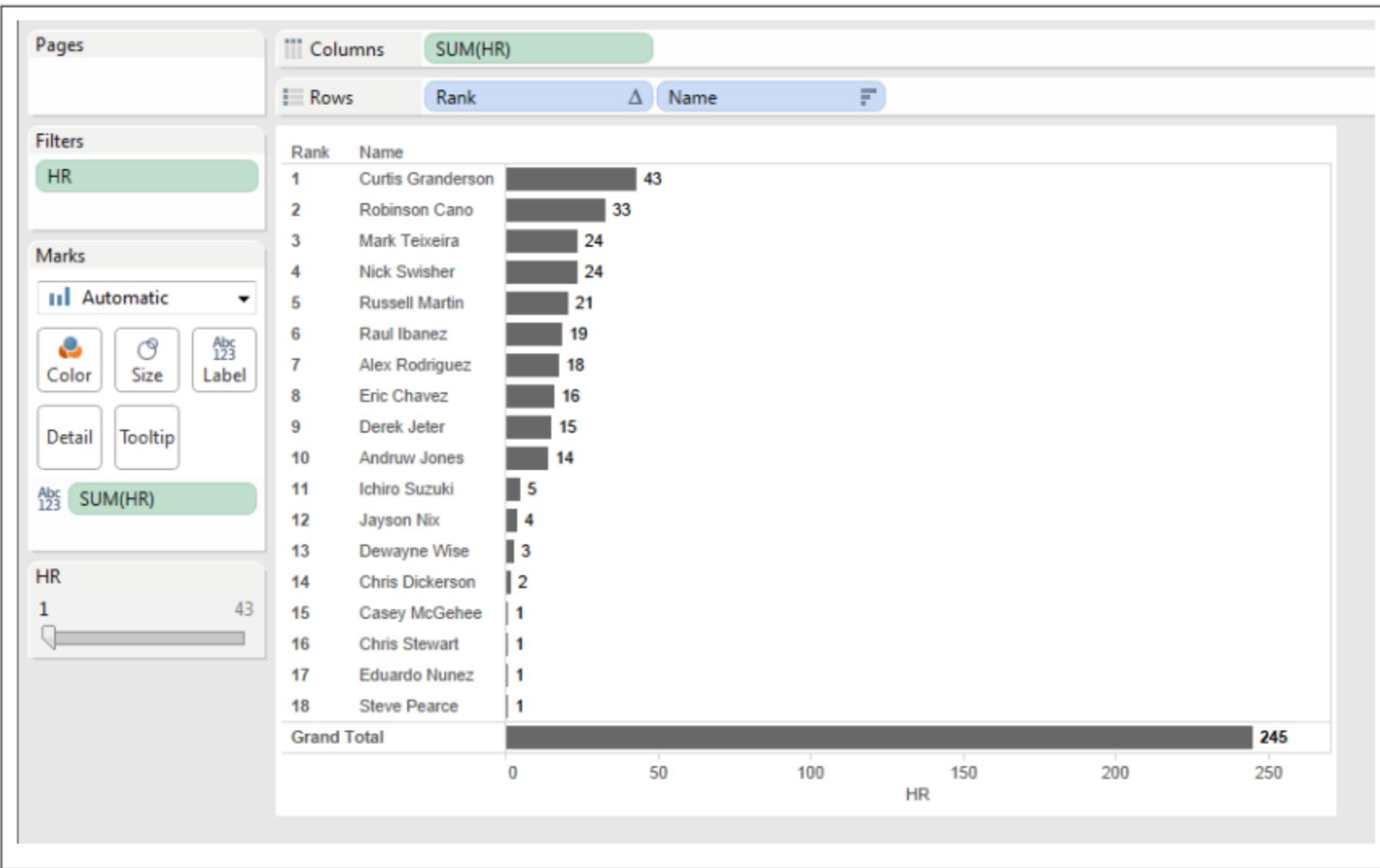


Figure 5-7. Home run bar chart with team total added

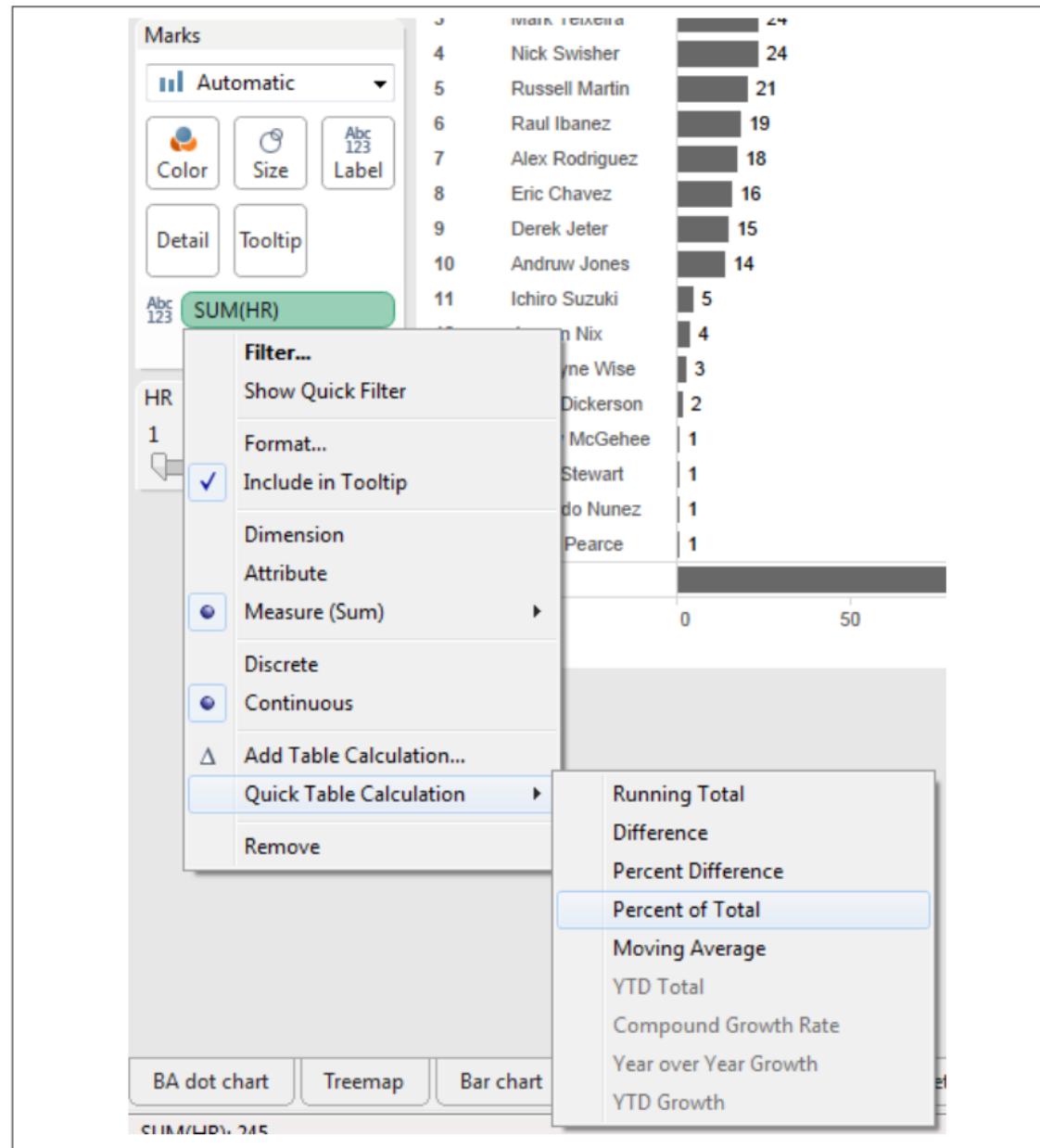


Figure 5-8. Changing the label to a “Percent of Total” table calculation

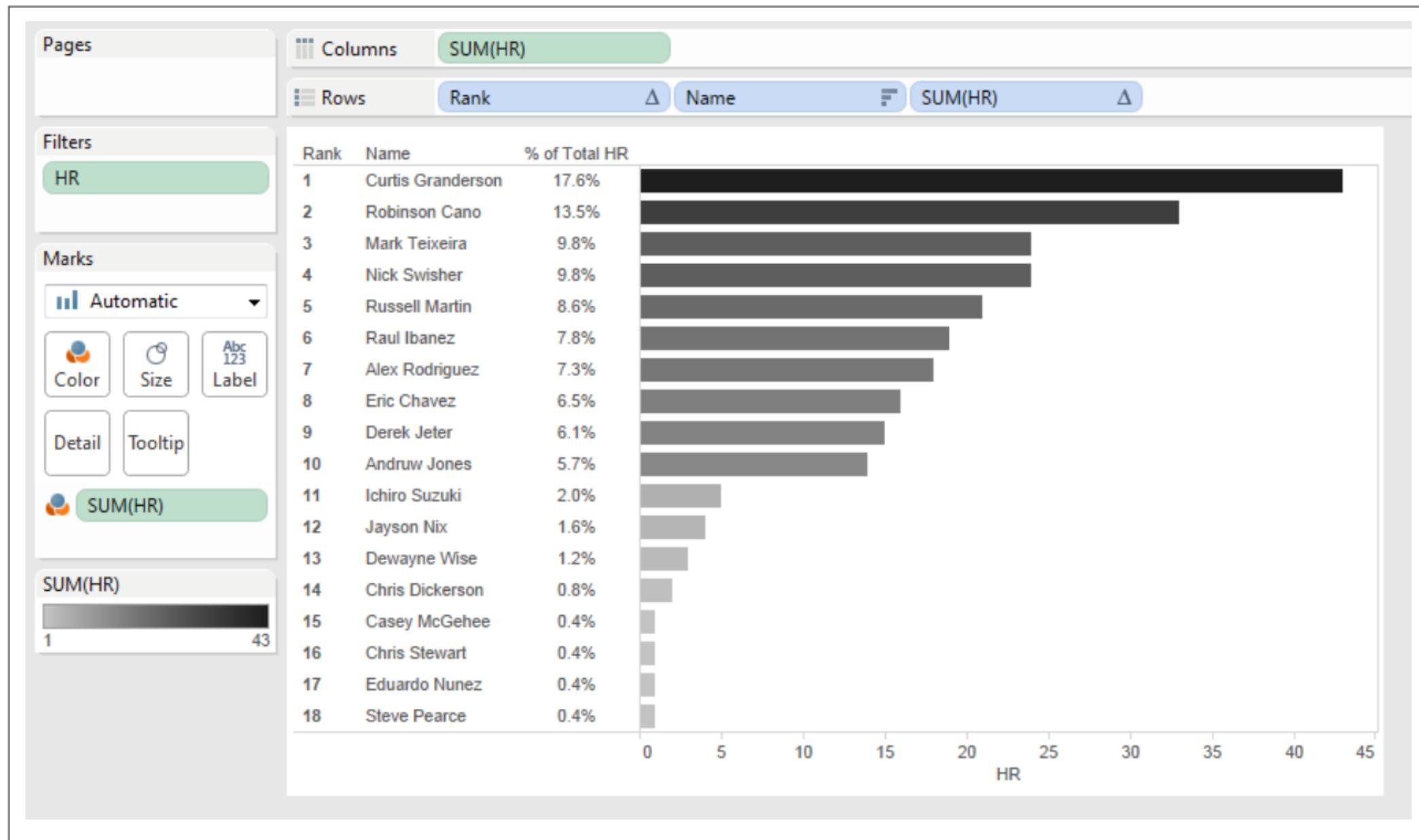
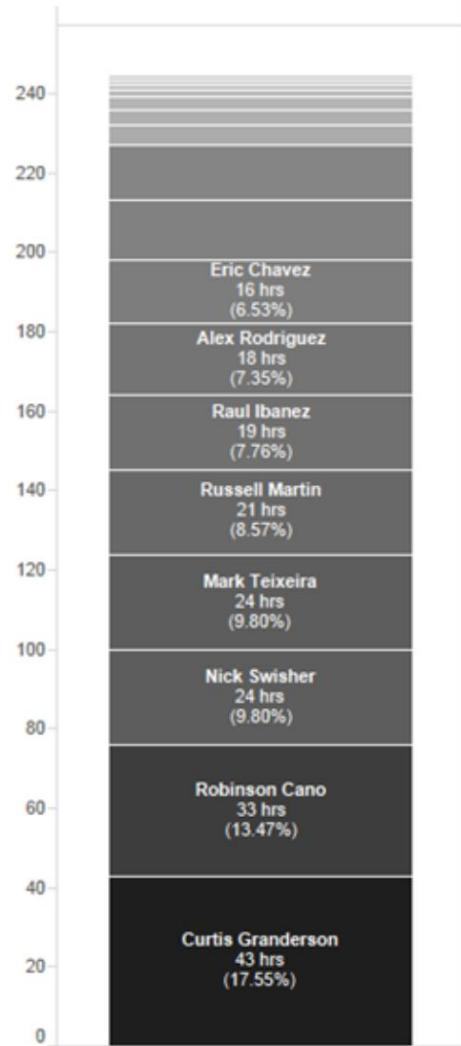
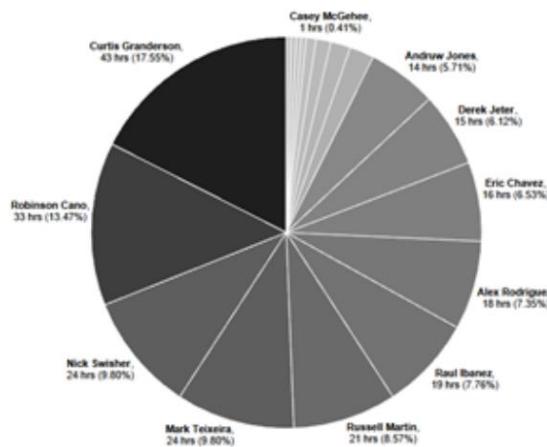


Figure 5-9. Home run tallies with labels shown as Percent of Total

1. Stacked bar



2. Pie chart



3. Treemap

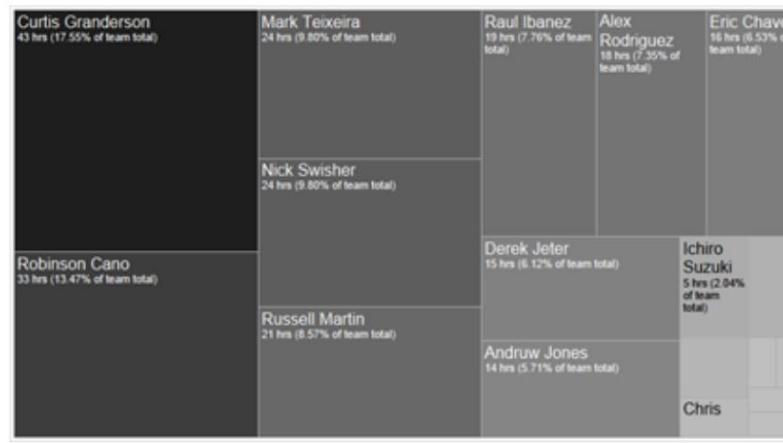
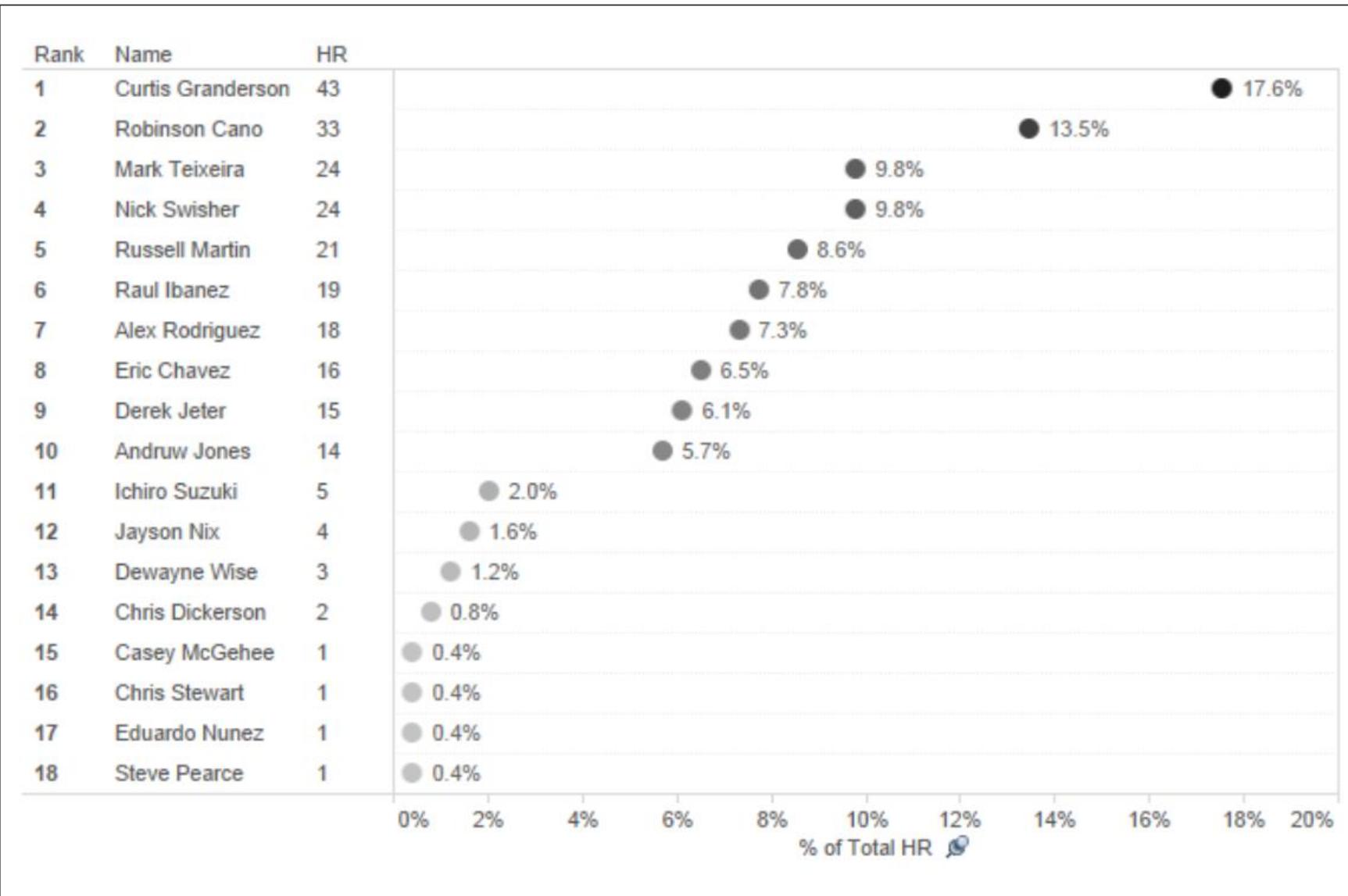
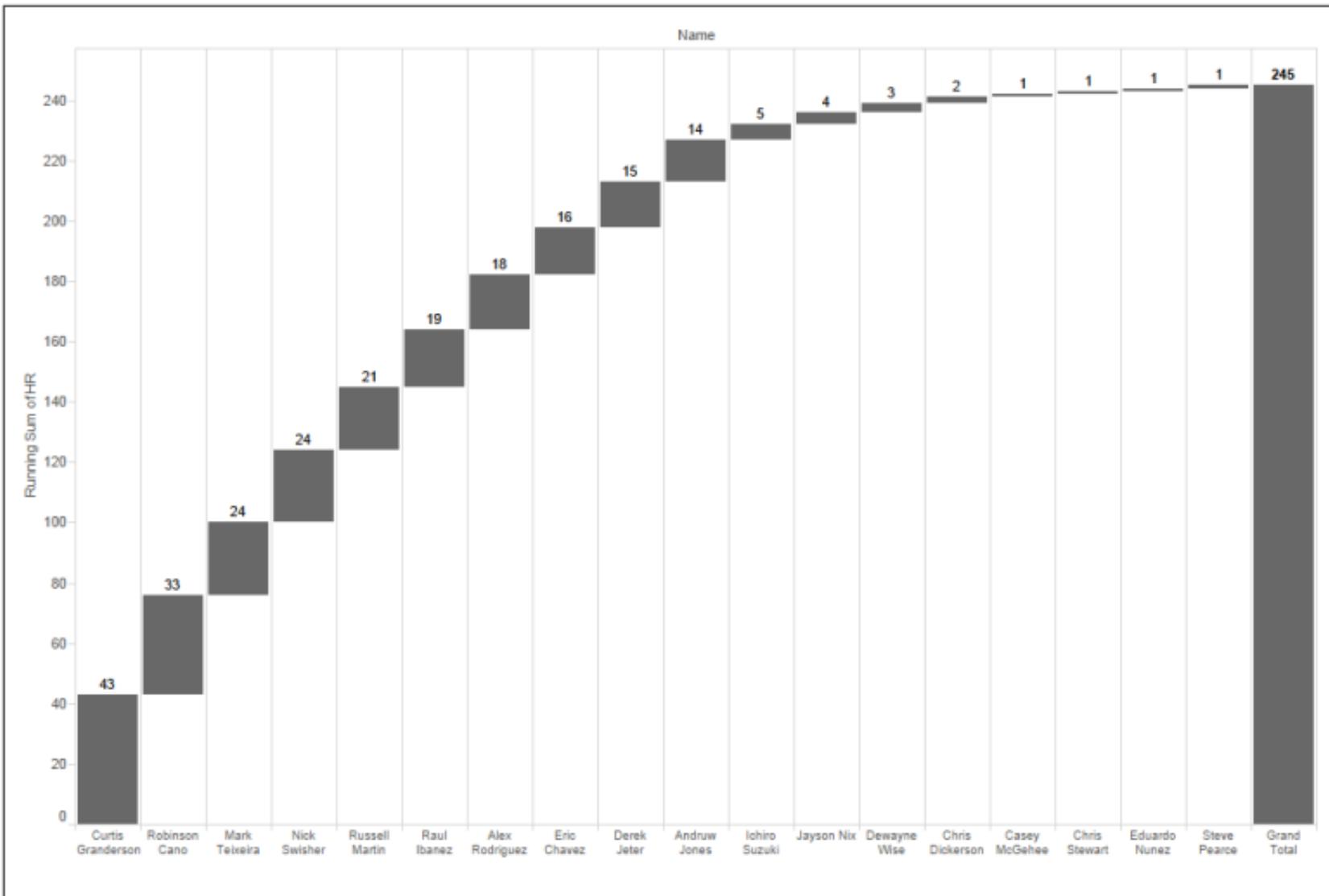


Figure 5-10. Three alternative ways to show the proportion of home runs hit by each player



*Figure 5-11. Dot chart of percentage of home runs contributed by each player*



*Figure 5-12. Home run data as a waterfall chart*



# Proportions and Percentages

1

Part to whole

2

Current to historical

3

Actual to target

4

Mean and Median



# Current to historical

- New York hit 245 home runs in 2012.
- But how did that home run tally compare with other teams in their league?
- how did it compare with the amount of home runs they hit the year before?

- Was it the most in the league, the least, or somewhere in between?
- Did they improve their home run tally over 2011 or did they fail to reach the previous year's mark?
- These questions send us in search of yet more data, and we find the 2011 and 2012 league team home run totals on the Web.

	A	B	C	D	E
1	Tm	City	League	2012 HR	2011 HR
2	BAL	Baltimore	AL	214	191
3	BOS	Boston	AL	165	203
4	CHW	Chicago	AL	211	154
5	CLE	Cleveland	AL	136	154
6	DET	Detroit	AL	163	169
7	KCR	Kansas City	AL	131	95
8	LAA	Los Angeles	AL	187	129
9	MIN	Minnesota	AL	131	103
10	NYY	New York	AL	245	222
11	OAK	Oakland	AL	195	114
12	SEA	Seattle	AL	149	109
13	TBR	Tampa Bay	AL	175	172
14	TEX	Texas	AL	200	210
15	TOR	Toronto	AL	198	186

Figure 5-15. Team home run totals, 2011 and 2012



# Current to historical

## The Bullet Graph:

- Let's go ahead and connect to the data table and create a bullet graph by Ctrl-selecting City, 2011 HR, and 2012 HR, and then choosing bullet graph from the Show Me panel. The resulting view is shown in Figure 5~16.

- In this initial view created from the Show Me menu, the length of the blue bars is determined by the 2011 HR totals, as evidenced by the green pill in the Columns shelf. The vertical black lines are the 2012 HR totals for each team, and the bands are at 60% and 80% of the 2012 values.

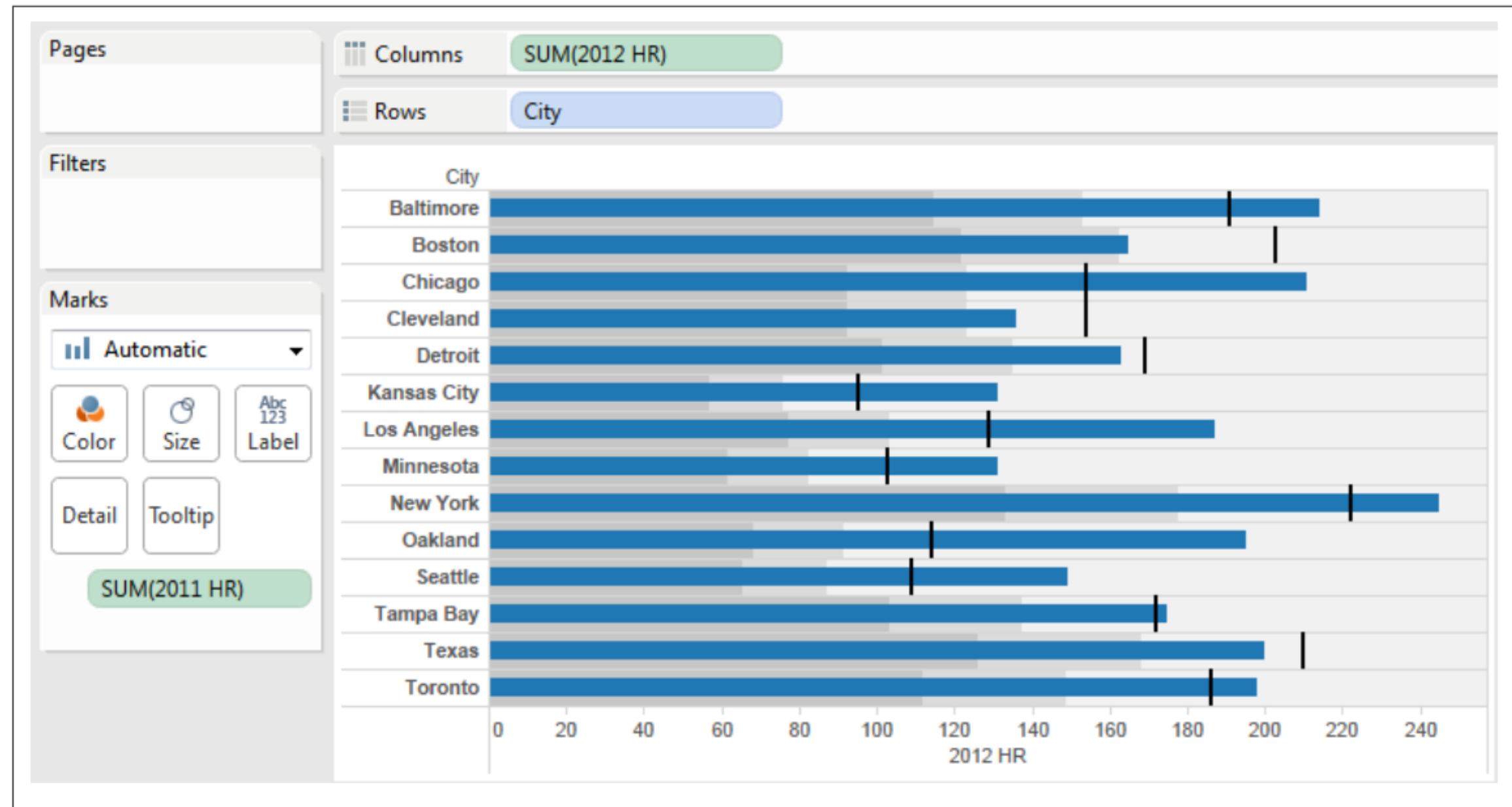


Figure 5-16. The initial bullet graph created by Show Me

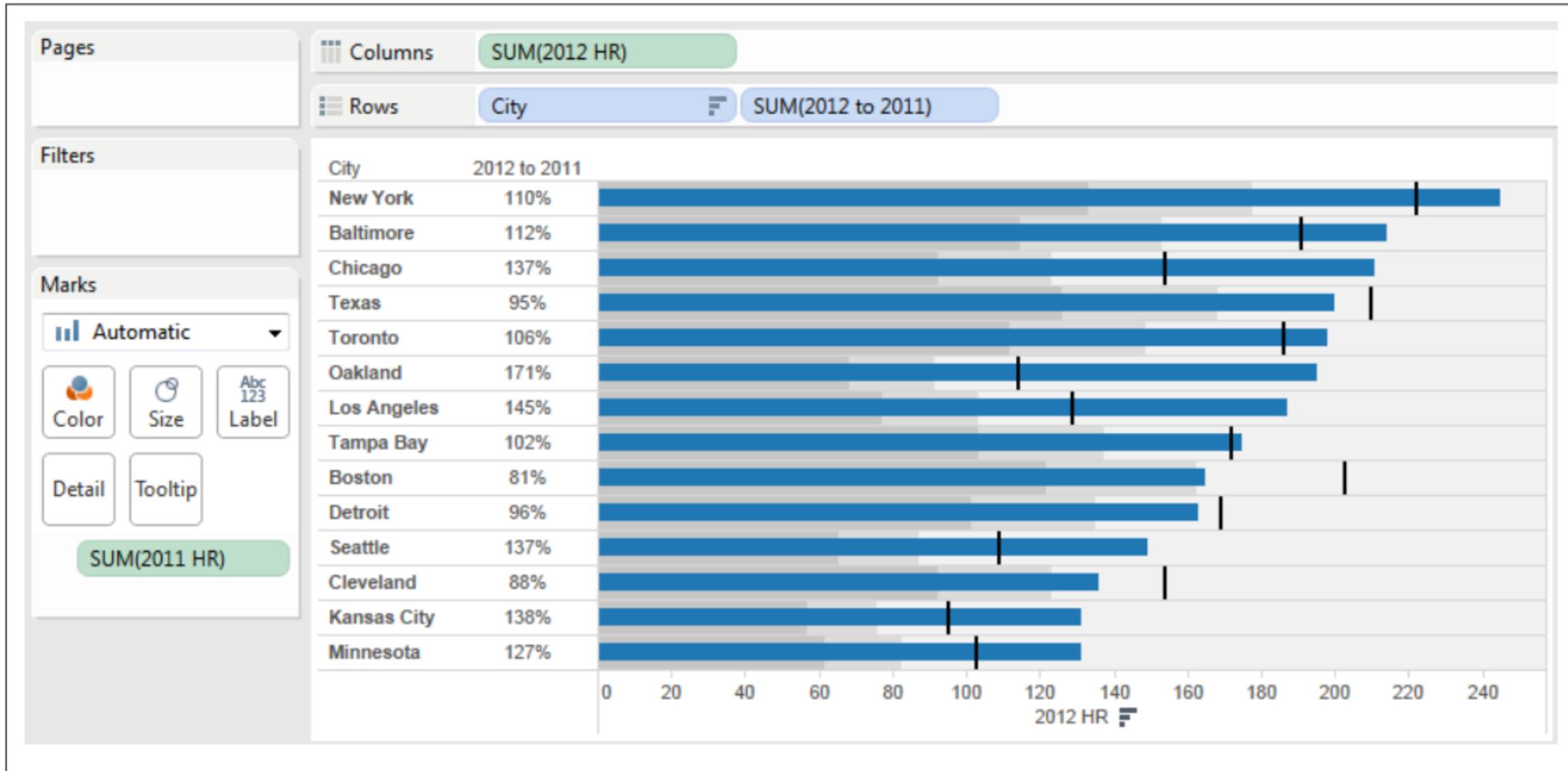


Figure 5-17. Bullet graph of AL team home runs, 2012 to 2011



# Current to historical

## How do we read the bullet graph?

- If the blue bar goes beyond the black line, then the team hit more home runs collectively in 2012 than they did in 2011.
- If the blue bar falls to the left of the black line, then the team hit fewer home runs, and their 2012 tally falls into one of three bands:

- The darkest gray band is for teams that hit less than 60% of their previous year total
- The next lighter gray band is for teams that hit between 60% and 80% of their previous year total
- The final band is for teams that hit between 80% and 100% of their previous year total. This time, we find



# Proportions and Percentages

1

Part to whole

2

Current to historical

3

Actual to target

4

Mean and Median



# Actual to target

- The world is full of people and teams with goals in the form of quotas, budgets, and performance targets. These figures are tracked and monitored religiously to determine “performance to plan.” Just listen to any monthly sales call.

- How are we doing compared to where we want to be at the end of the year?
- How are we doing compared to where we should be right now if we want to hit our monthly goal?

## CHARTING ACTUAL TO TARGET VALUES

Monthly Target

Mtd Actual

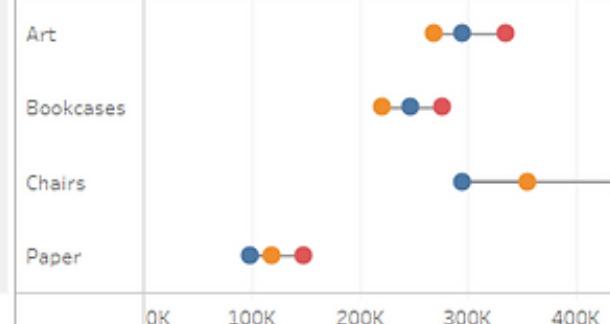
Mtd Target

where Mtd = Month-to-date

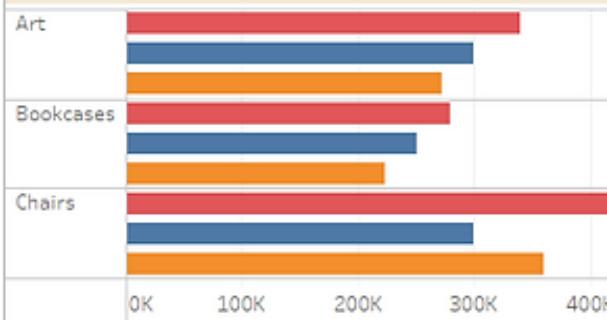
### Bullet chart



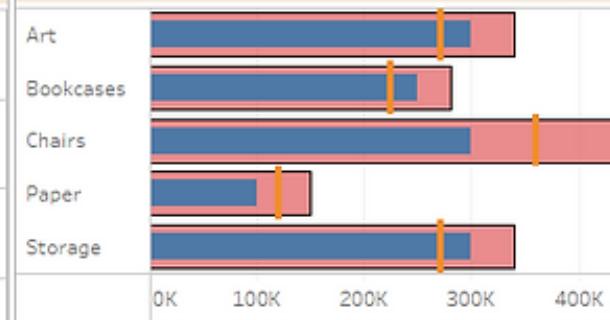
### Dot plot



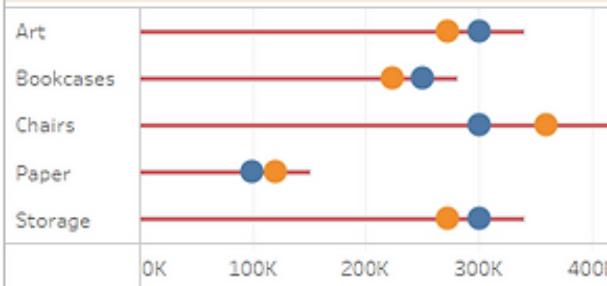
### Side-by-side bars



### Bar-in-bar chart + Reference line



### Bar + dot plot



### BANS

340,000

▲10.3% Above Mtd target

▼-11.8% Below Monthly Target



# Proportions and Percentages

1

Part to whole

2

Current to historical

3

Actual to target

4

Mean and Median



# Mean and Median

- The **mean (or average)** is determined by summing all the values in a data set and dividing by the number of values.
- The mean is considered a “representative value,” meaning if you replaced each value in the data set with the mean, the overall sum wouldn’t change.
- The **median** is the middle value in a data set in which the values have been placed in order of magnitude. Thus, half the values in the data set are less than the median, and half are greater.
- The **mode** is the most commonly occurring value in a data set.

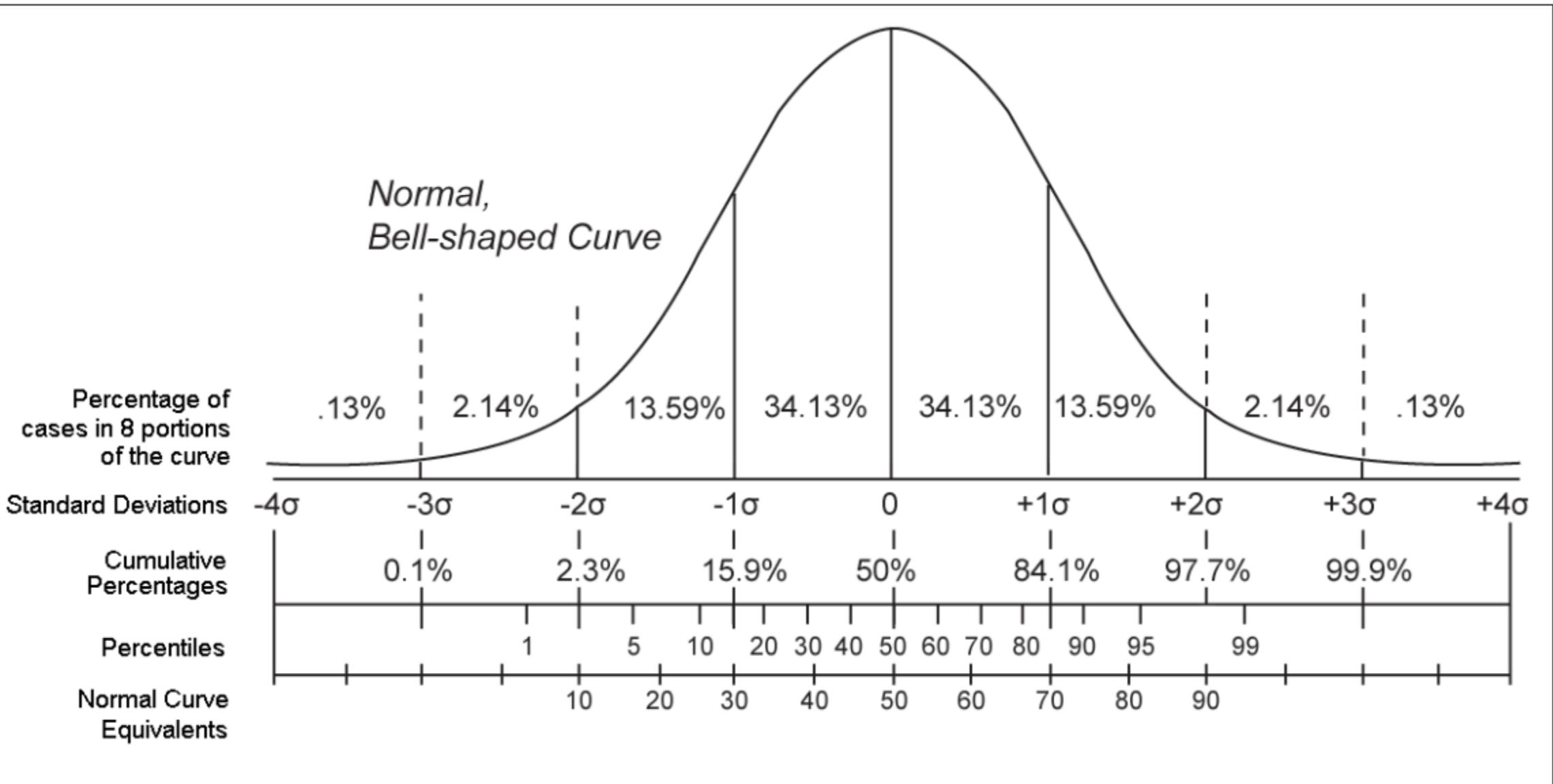


# Mean and Median

## The Normal Distribution:

- The Gaussian, or normal, distribution, is something we've all been exposed to at some point.

- The unmistakable bell-shaped curve of the Gaussian represents a very mild style of variation, one in which the probability of a value occurring falls off dramatically the farther we move away from the mean on either side



*Figure 6-1. The characteristics of the **normal distribution***

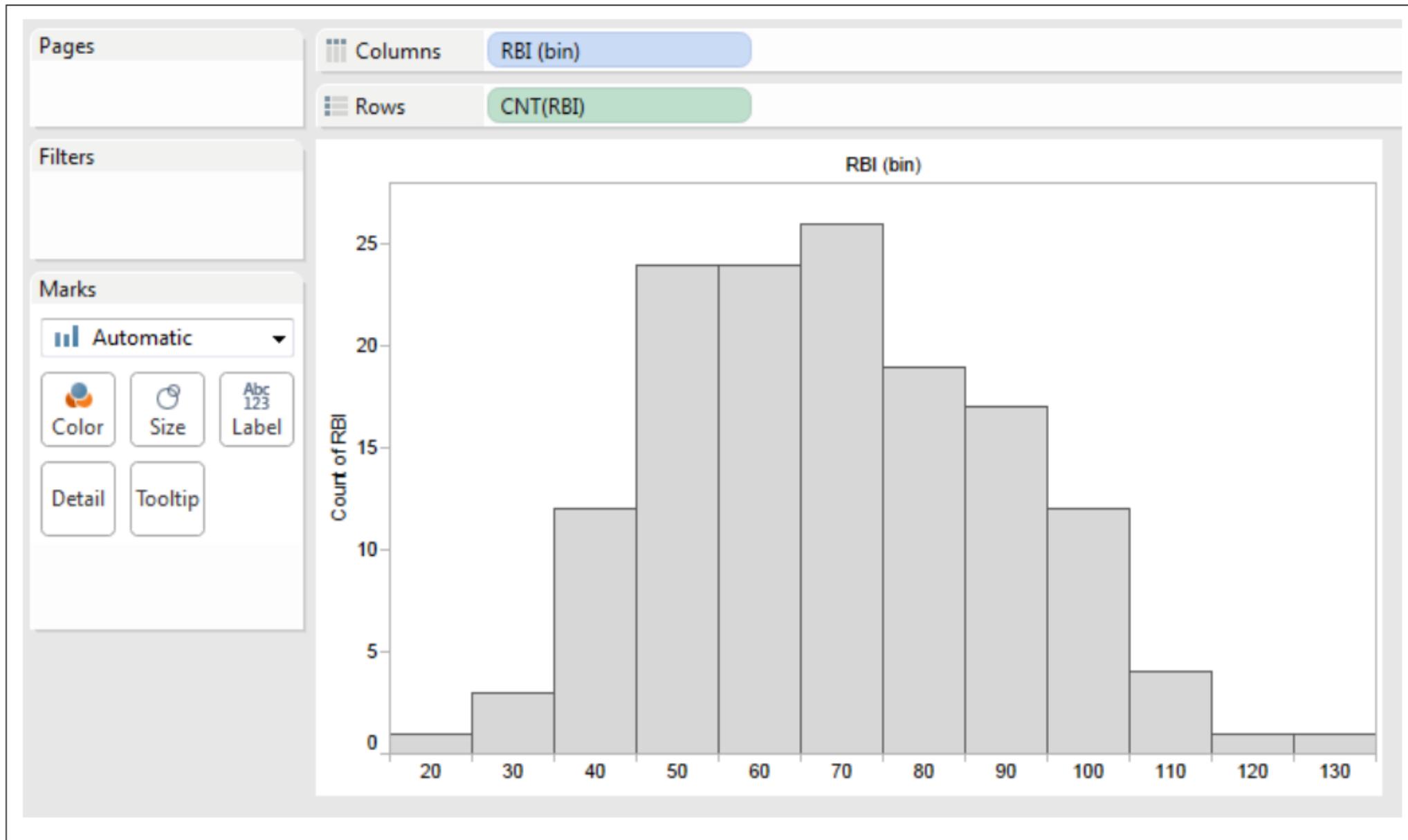


# Mean and Median

## An Example of “Normal” Data:

- In baseball, a run batted in (RBI) is granted to a batter every time he enables a runner to score during his at bat. A batter can earn more than one RBI during a single at bat; a grand slam home run would result in 4 RBI

- We can create a histogram as before, to visualize the distribution of qualifying players’ RBI during the 2012 season, as shown in Figure 6~2.



*Figure 6-2. A histogram of players' RBI during the 2012 season*



# Mean and Median

## Box Plots:

- If we Ctrl-click the Player and Pos (for “Position”) Dimensions, and the RBI Measure, and then open the Show Me panel and select box-and-whisker plot, we get the chart shown in Figure 6~4.

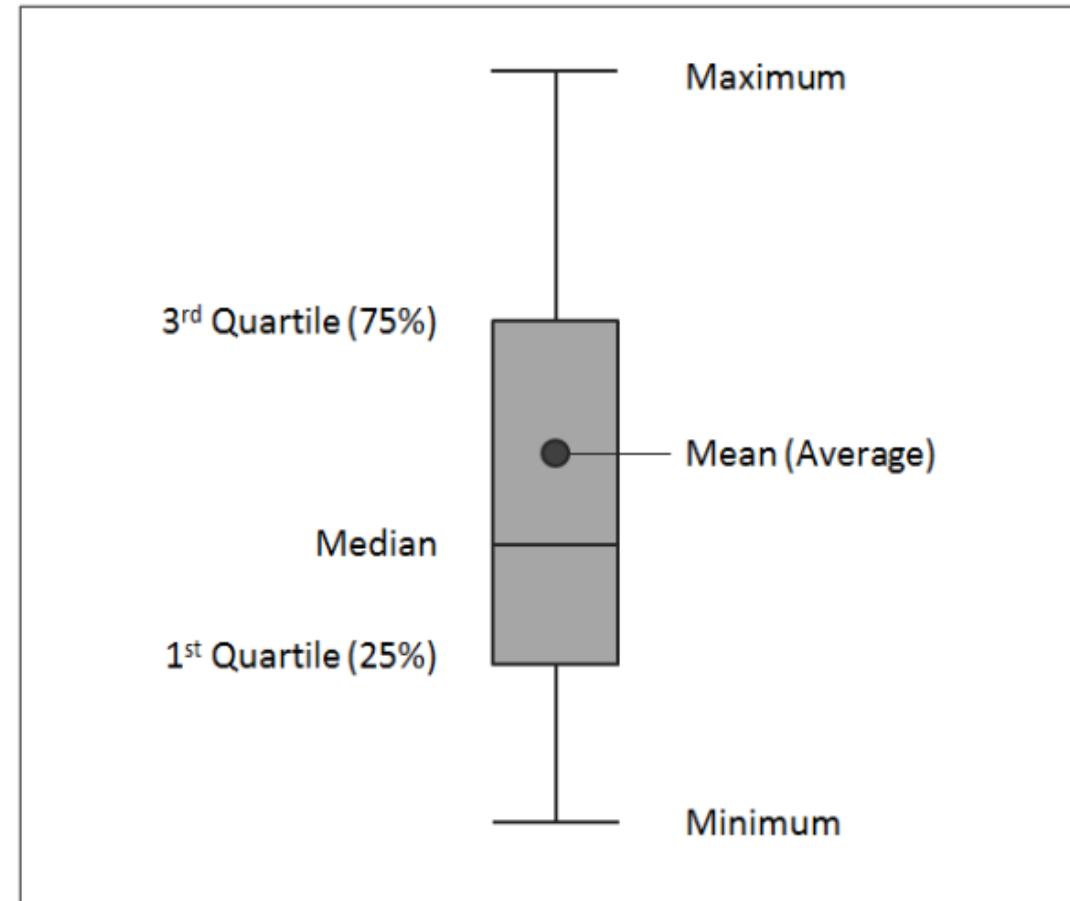


Figure 6-3. The box plot

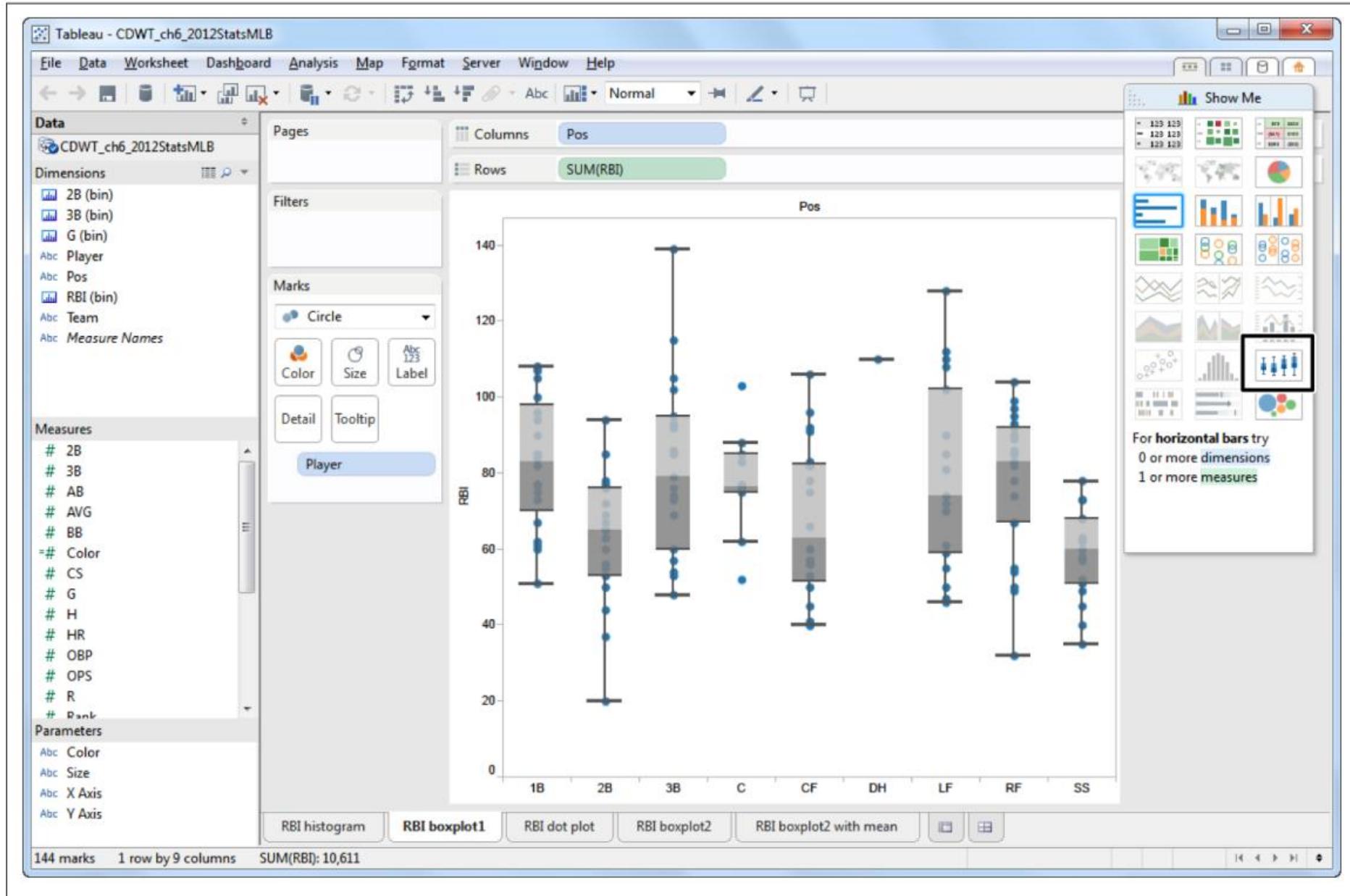


Figure 6-4. The box-and-whisker plot

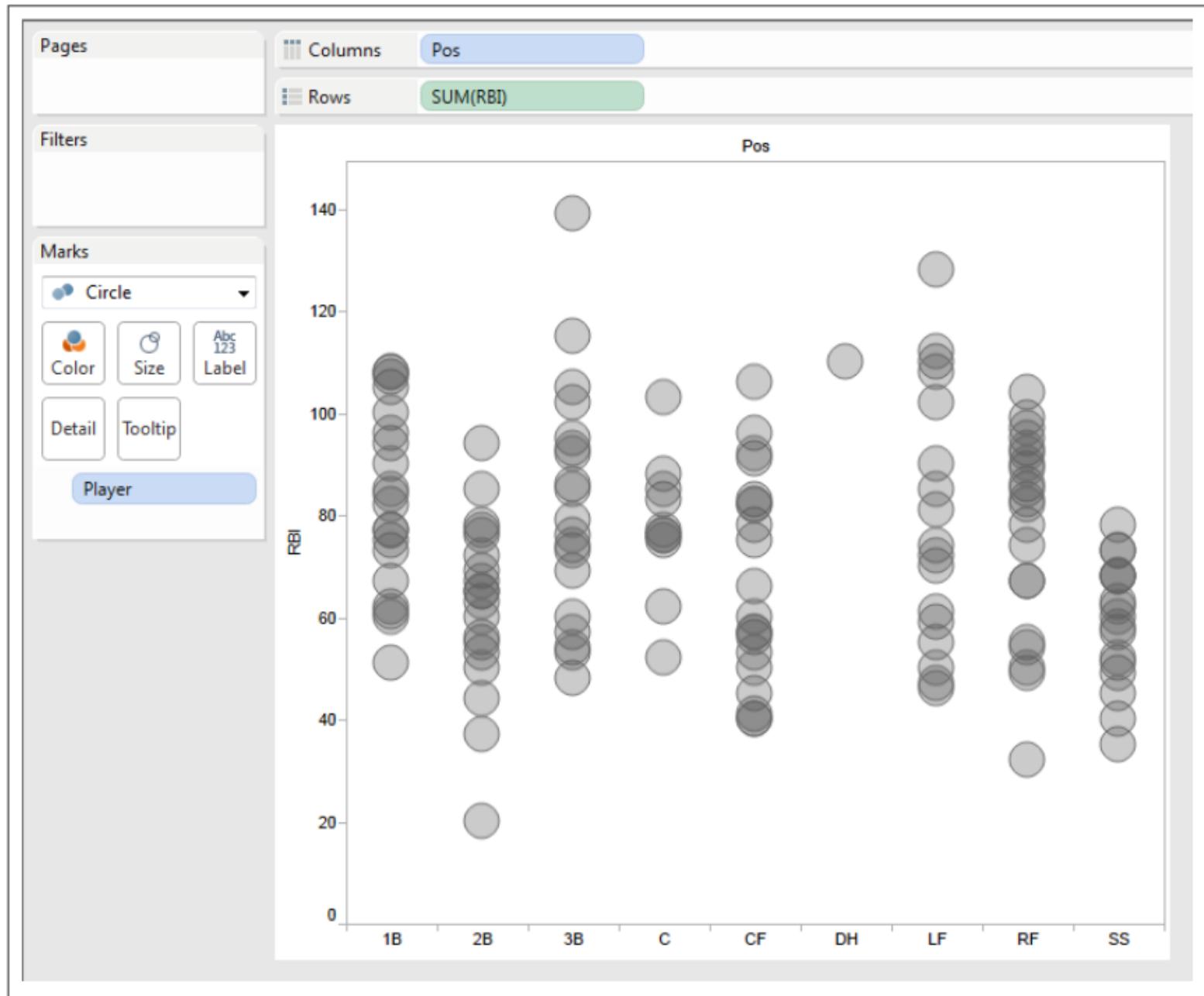


Figure 6-5. Dot plot of RBI by position

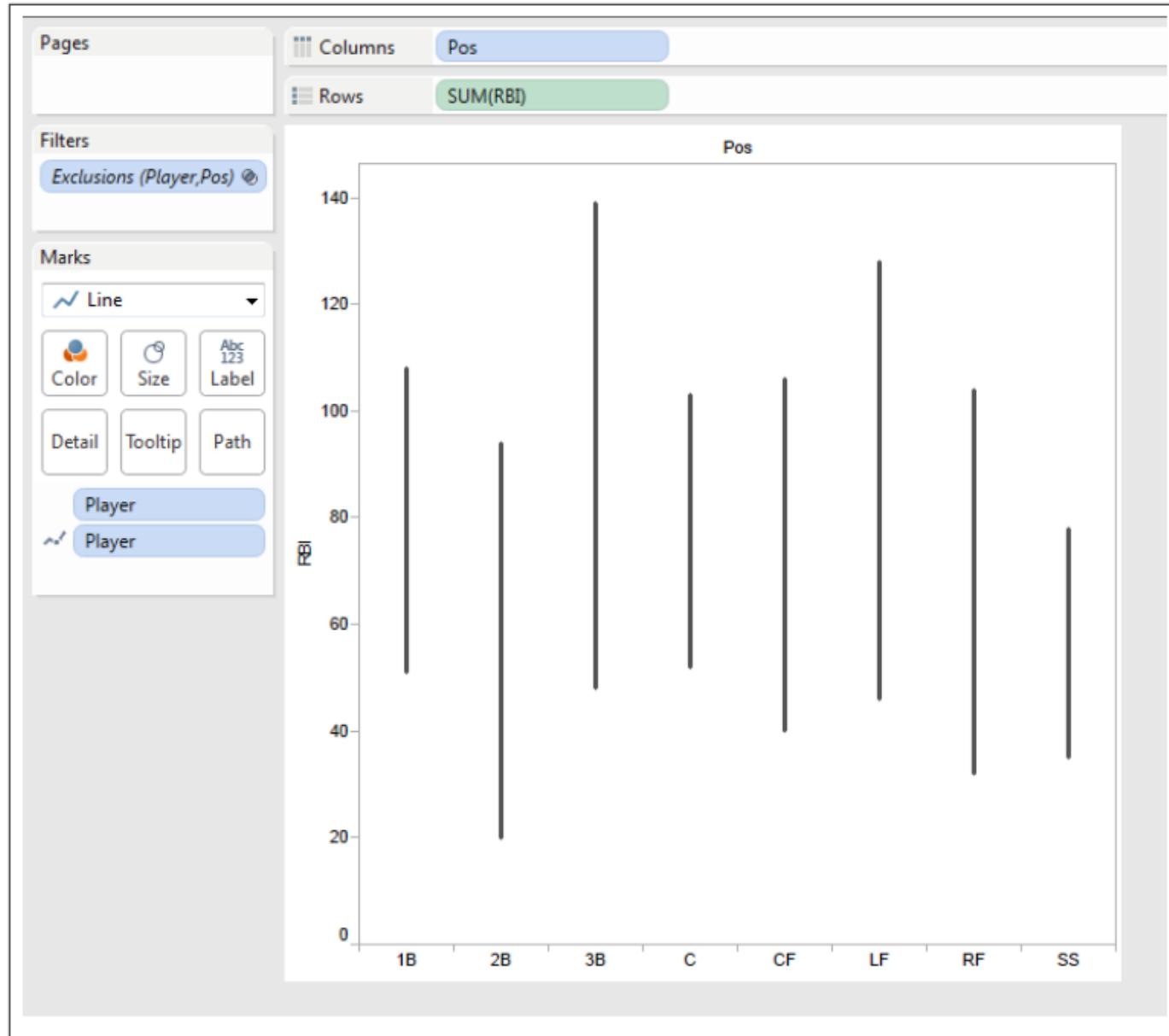


Figure 6-6. Vertical lines from the min to the max values of each position



# DATA VISUALIZATION



# **UNIT - IV (PART - II)**



# Variation and Uncertainty

- It's good to feel confident when armed with a bit of data, but it's crucial to stay humble.
- The world is complex and ever-changing.
- If our data is unreliable or our conclusions are questionable, we should be cautious.
- When sharing data, honesty is key.
- Clearly communicate what we know, what we don't, and represent reality as accurately as possible.
- If data has high variation or is from a limited sample, be transparent to avoid misleading our audience.



# Variation and Uncertainty

- Variation refers to how much individual observations differ within a group.
- Example: Students in a class may have different heights because of factors like genetics and nutrition.

- Uncertainty is the lack of confidence in making inferences about a population from data collected in samples.
- Example: It's hard to be completely sure about the average income of a city when relying on a small survey.



# Variation and Uncertainty

1

Respecting variation

2

Variation over time-Control charts

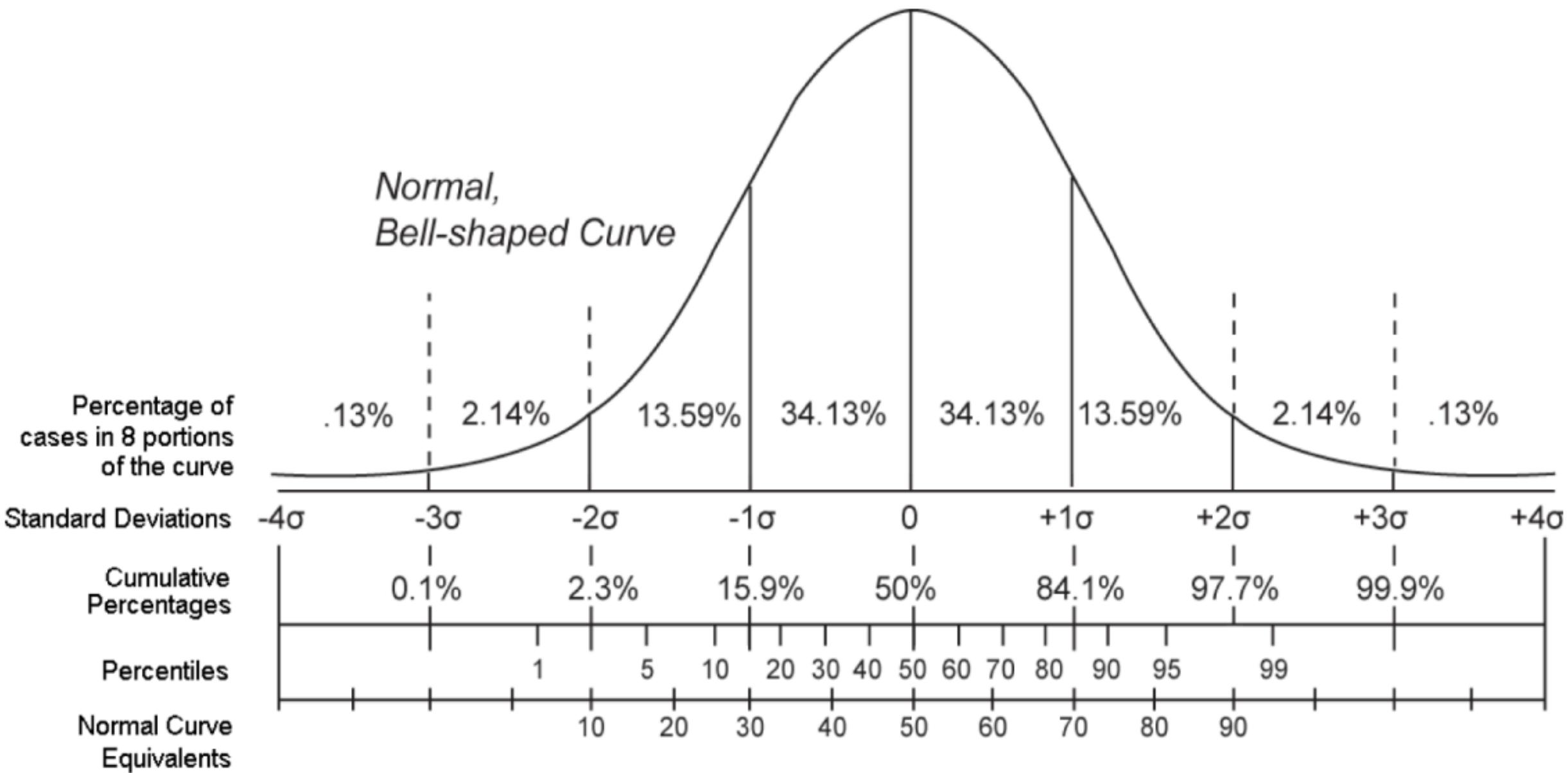
3

Understanding uncertainty



# Respecting variation

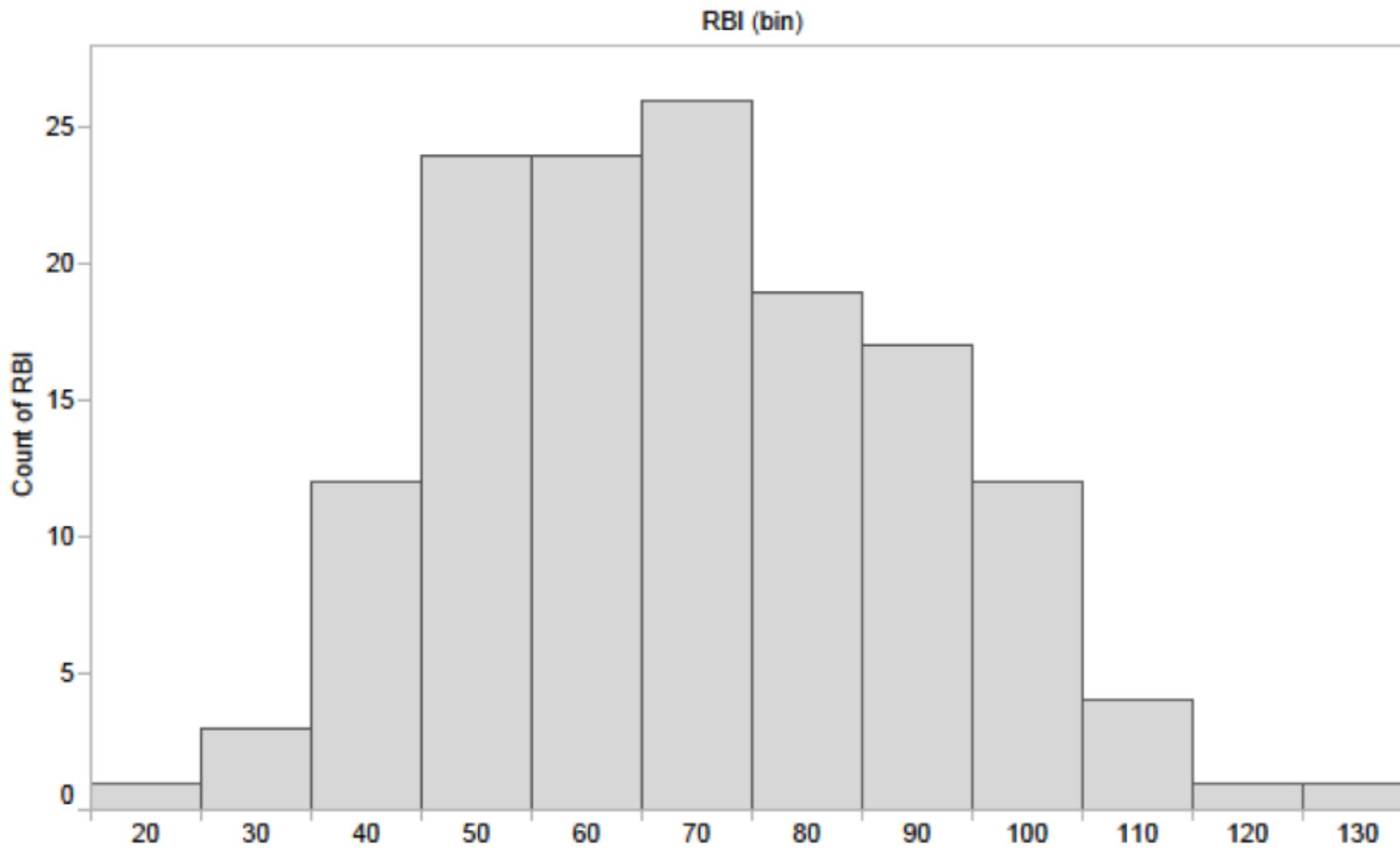
- In the previous chapter, we explored central tendency measures, such as the mean and median.
- This discussion also included fundamental measures of variation, like standard deviation and the interquartile range, as illustrated in Figure 7~1.





# Respecting variation

- In the previous chapter, we examined two distinct types of variables in the realm of sports: baseball batting statistics (RBI) and soccer players' salaries, depicted in Figure 7~2.



2012 Guaranteed Compensation (bin)



Thierry Henry (NY): \$5.6M in  
guaranteed compensation

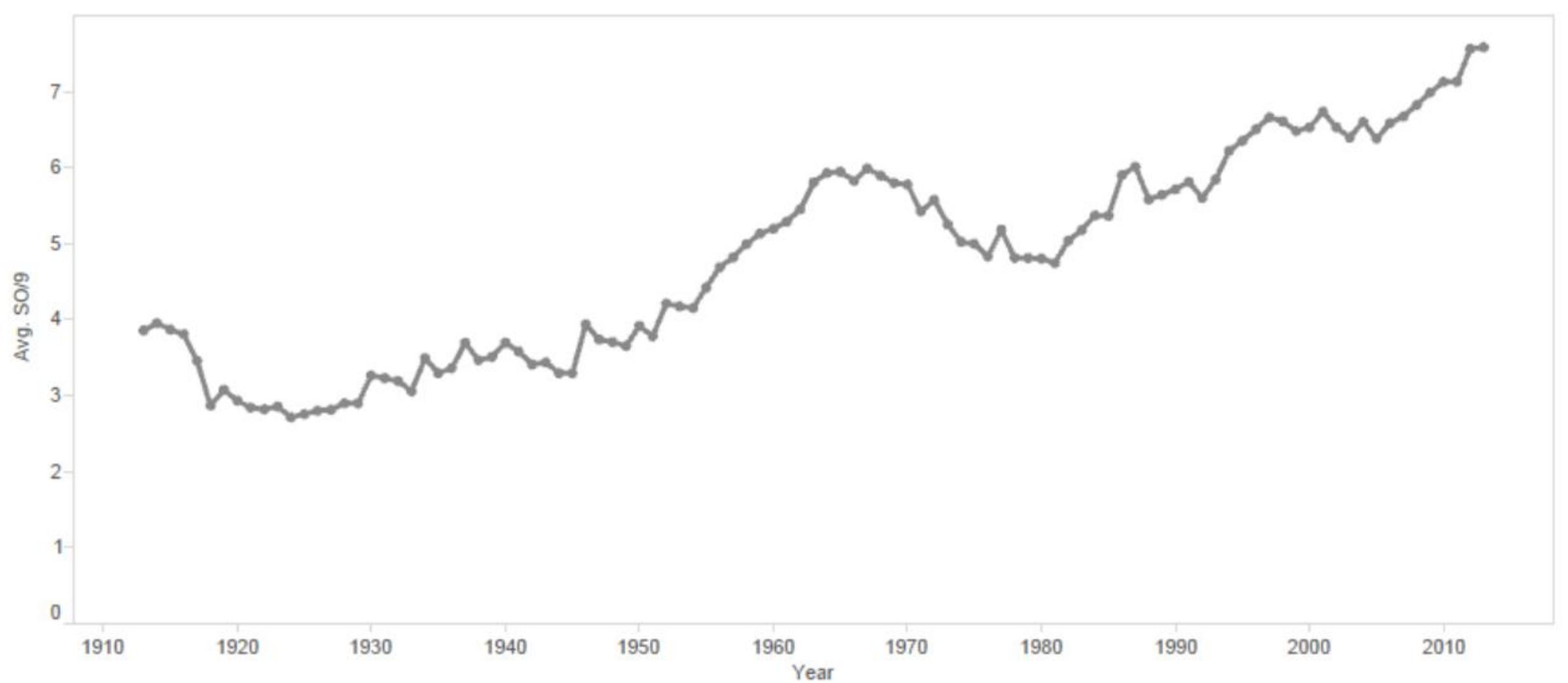


# Respecting variation

## Visualizing Variation:

- To honor the inherent variation in our data, it's essential to display it.
- Presenting only averages creates a too simple view of the world.
- Just as not every person in a country shares the most common physical traits, not every data point aligns with the mean, median, or mode.

- If we consider once again the number of strikeouts per nine innings in professional baseball over the past 100 years, we can show a simple line plot of average strikeouts per nine innings, as shown in Figure 7~3



*Figure 7-3. Average number of strikeouts per nine innings*

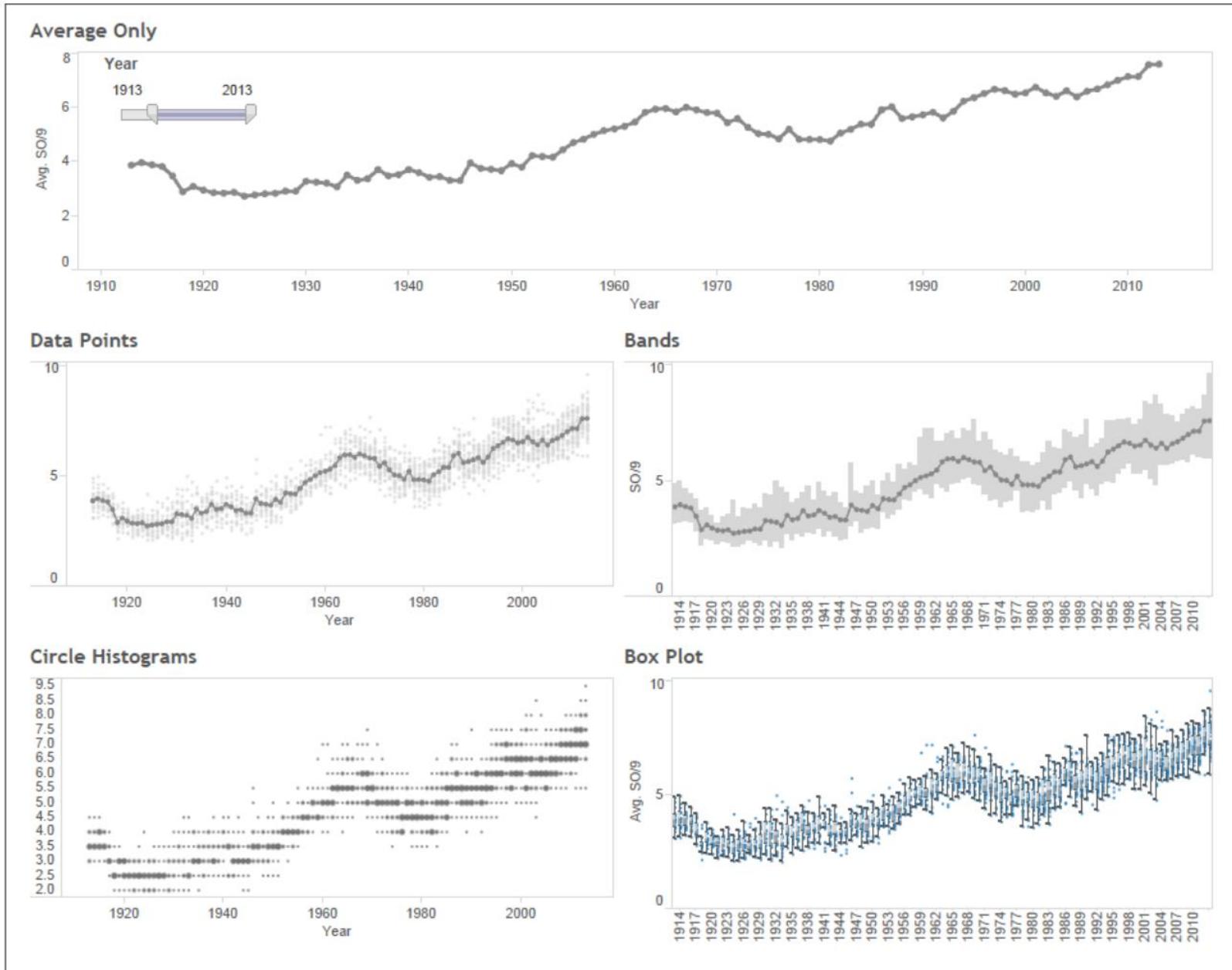


# Respecting variation

## Visualizing Variation:

- However, this chart doesn't reveal how the strikeout rates varied among different teams in the league each year.
- We're left in the dark about the contrast between the team with the highest strikeout rate and the one with the lowest rate annually.

- To capture the inherent variation in the data, we can represent it in various ways, as depicted in Figure 7~4



*Figure 7-4. Four different ways of showing variation in a time series*



# Respecting variation

## Data points:

- Each team is represented by its own circle in each year.

## Bands:

- Includes reference bands from the minimum to the maximum for each year.

## Circle Histograms:

- Consists of circle histograms, where the area of each circle is proportional to the number of teams in each bin.

## Box~Plots

- Displays a series of box plots for each year.



# Variation and Uncertainty

1

Respecting variation

2

Variation over time-Control charts

3

Understanding uncertainty



# Variation over time- Control charts

- Control charts help determine if data collected over time contains statistically significant signals or if the variation is just noise.
- Walter Shewhart developed them in the 1920s at the Western Electric Company for industrial quality control.
- The Six Sigma movement has popularized these charts, with "black belts" using them to measure process behavior and reduce variation to enhance quality.



# Variation over time-Control charts

- The idea is that reducing variation leads to fewer defects.
- This concept is particularly applicable in manufacturing and any scenario where a consistent output is essential.
- For instance, when ordering a burger from a fast-food chain or starting a new car, we expect a standardized product.
- In such cases, variation would likely be undesirable.

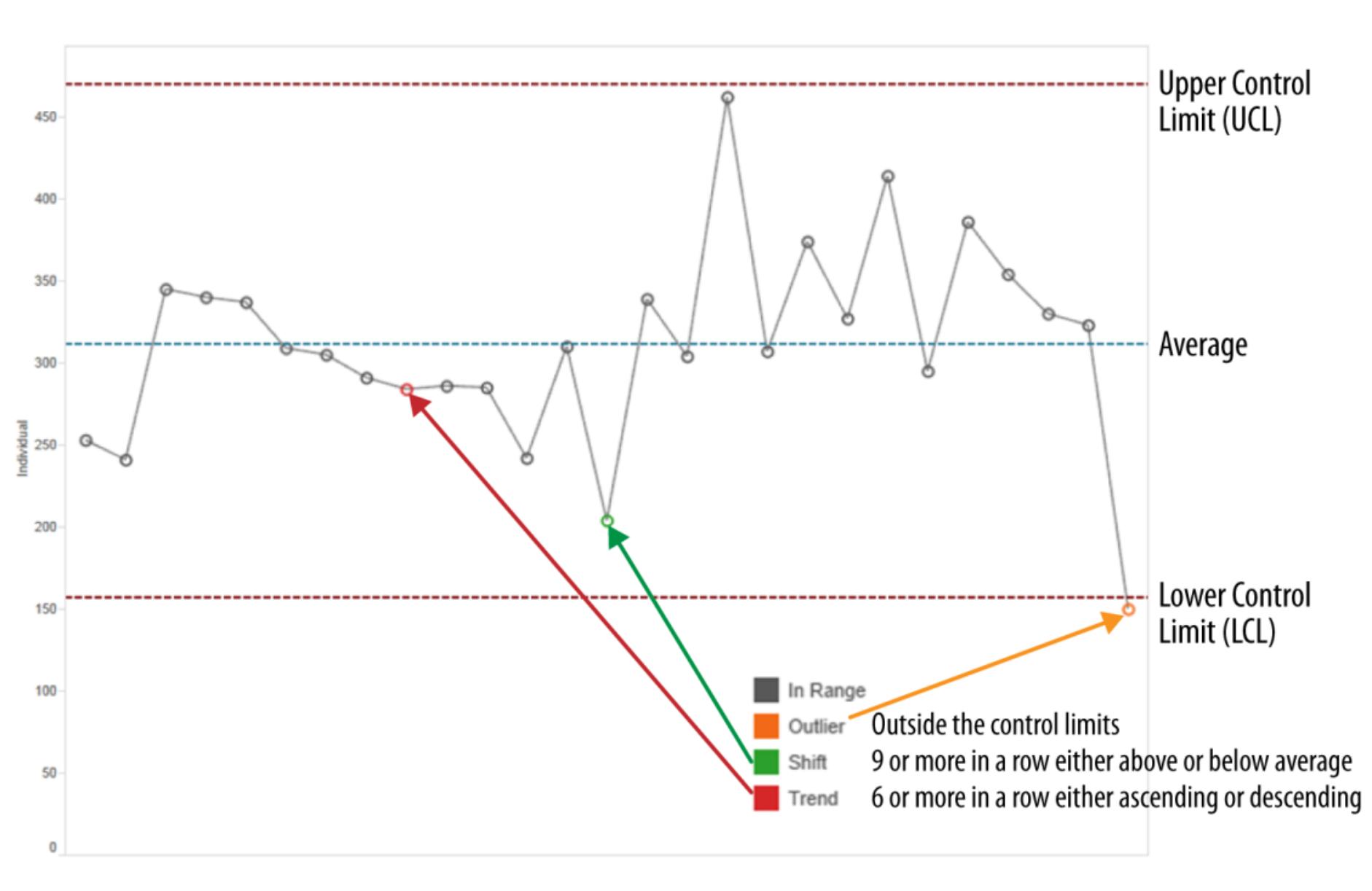


Figure 7-5. The elements of a Shewhart Control Chart



# Variation over time-Control charts

## Anatomy of a Control Chart:

- A control chart contains the following basic elements:
  1. The time series data itself
  2. The average line
  3. The control limits: UCL (the upper control limit) LCL (the lower control limit)

## 4. Signals:

- Outliers (data points either above the UCL or below the LCL)
- Trends (six or more points either all ascending or all descending)
- Shifts (nine or more points either all above or all below the average line)



# Variation over time-Control charts

## How to Create a Control Chart in Tableau

- Let's explore two methods for control chart analysis: the quick method and the rigorous method.
- The main distinction lies in how the control limits are determined.

- The **quick method** employs a global measure of dispersion, specifically the standard deviation of all data points.
- The **rigorous method** utilizes a local measure of dispersion known as  $\Sigma(x)$ , derived from the differences between successive data points.



# Variation over time-Control charts

## Example:

- Consider the total number of earthquakes recorded worldwide that registered magnitude 6.0 or higher on the Richter scale from 1983 through 2013.

- The source for the data is the USGS Earthquake Archive Search website.
- There were 4,136 such events recorded, and Figure 7-6 gives a view of the most recent records in the data set

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date & Time	latitude	longitude	depth	mag	magType	nst	gap	dmin	rms	net	id	updated	place	type
2	2013-12-17 23:38:06	20.7727	146.7903	9	6.2	mww		14	4.162	0.81	us	usc000lmmc	2014-02-27T21:49:52.353Z	198km E of Farallon de Pajaros, Northern Mariana Islands	earthquake
3	2013-12-08 17:24:54	44.4438	149.1667	28	6	mww		26	4.726	0.99	us	usb000lds9	2014-02-15T02:35:45.869Z	134km SE of Kuril'sk, Russia	earthquake
4	2013-12-01 06:29:57	2.044	96.8261	20	6	mww		27	1.049	0.89	us	usb000l8pb	2014-02-12T02:20:23.821Z	69km SE of Sinabang, Indonesia	earthquake
5	2013-12-01 01:24:13	-7.0269	128.3791	9.87	6.4	mww		11	3.046	0.7	us	usb000l8mb	2014-02-12T02:20:47.278Z	Kepulauan Barat Daya, Indonesia	earthquake
6	2013-11-25 07:21:18	-53.8708	-53.9107	14.83	6	mwc		58	3.346	0.76	us	usb000l633	2014-02-11T02:22:33.206Z	South Atlantic Ocean	earthquake
7	2013-11-25 06:27:33	-53.9451	-55.0033	11.78	7	mww		31	2.935	1.08	us	usb000l5zn	2014-02-11T02:25:27.101Z	Falkland Islands region	earthquake
8	2013-11-25 05:56:50	45.5613	151.0047	34	6	mww		26	5.885	0.66	us	usb000l5z1	2014-02-11T02:32:38.383Z	247km E of Kuril'sk, Russia	earthquake
9	2013-11-23 07:48:32	-17.1171	-176.5449	371	6.5	mww		22	5.194	0.83	us	usb000l51g	2014-02-11T02:22:07.739Z	Fiji region	earthquake
10	2013-11-19 17:00:44	18.4753	145.2041	511	6	mww		10	1.848	1.05	us	usb000l25i	2014-02-11T02:29:01.431Z	58km WSW of Agrihan, Northern Mariana Islands	earthquake
11	2013-11-19 13:32:51	2.6403	128.4339	38	6	mww		19	2.14	1.01	us	usb000l219	2014-02-11T02:36:43.624Z	111km NNE of Tobelo, Indonesia	earthquake
12	2013-11-17 09:04:55	-60.2738	-46.4011	10	7.7	mww		23	8.05	1.33	us	usb000l0gq	2014-01-31T21:29:01.439Z	Scotia Sea	earthquake
13	2013-11-16 03:34:31	-60.2627	-47.0621	9.97	6.9	mww		17	8.284	0.84	us	usb000kznc	2014-01-31T21:32:02.803Z	Scotia Sea	earthquake
14	2013-11-13 23:45:47	-60.2814	-47.1233	11.07	6.1	mww		23	8.319	1.19	us	usb000kxhr	2014-01-31T21:26:05.362Z	Scotia Sea	earthquake
15	2013-11-12 07:03:51	54.6859	162.3024	43	6.4	mww		20	2.73	0.87	us	usb000kw1x	2014-01-31T21:35:40.466Z	172km S of Ust'-Kamchatsk Staryy, Russia	earthquake
16	2013-11-02 18:53:46	-19.1711	-172.6411	10.05	6.2	mww		21	5.297	0.72	us	usb000krlz	2014-01-10T13:04:16.196Z	152km ESE of Neiafu, Tonga	earthquake
17	2013-11-02 15:52:46	-23.6357	-112.5956	9.98	6	mww		35	4.558	0.81	us	usb000krjt	2014-01-10T13:03:30.639Z	Easter Island region	earthquake
18	2013-10-31 23:03:59	-30.2921	-71.5215	27	6.6	mww		31	0.636	1.28	us	usb000kqnc	2014-01-10T13:06:12.899Z	41km SSW of Coquimbo, Chile	earthquake
19	2013-10-31 12:02:08	23.5904	121.4366	10	6.3	mww		15	0.234	1.29	us	usc000ksdy	2014-01-10T13:05:25.588Z	46km SSW of Hualian, Taiwan	earthquake
20	2013-10-30 02:51:47	-35.314	-73.395	41.5	6.2	mww			1.68	us	usc000kr9k	2014-01-10T13:04:41.053Z	88km W of Constitucion, Chile	earthquake	
21	2013-10-25 17:10:19	37.1557	144.6611	35	7.1	mww		10	3.968	1.01	us	usc000kn4n	2014-01-03T00:48:15.801Z	Off the east coast of Honshu, Japan	earthquake
22	2013-10-24 19:25:10	-58.153	-12.7964	22.87	6.7	mww		53	13.711	0.99	us	usc000kmfw	2014-01-03T00:40:15.133Z	East of the South Sandwich Islands	earthquake
23	2013-10-23 08:23:30	-23.0067	-177.1425	160	6	mwb		19	6.252	0.84	us	usb000kj1z	2014-02-21T19:59:38.000Z	283km SW of Vaini, Tonga	earthquake

Figure 7-6. Sample of global earthquakes data set, registering magnitude 6.0 or greater



# Variation over time-Control charts

## Quick Method:

- Create a simple timeline with a YEAR (Date & Time) on the Columns shelf, and SUM (Number of Records) on the Rows shelf, fit to width as shown in Figure 7~7.

- Right-click on the y-axis, select Add Reference Line, and add an average line by filling out the resulting dialog box.
- Then right-click on the y-axis, select Add Reference Line again, and this time add a distribution of +3 and -3 times the standard deviation, with dotted red lines and no fill.
- Both reference line dialog boxes are shown in Figure 7~8

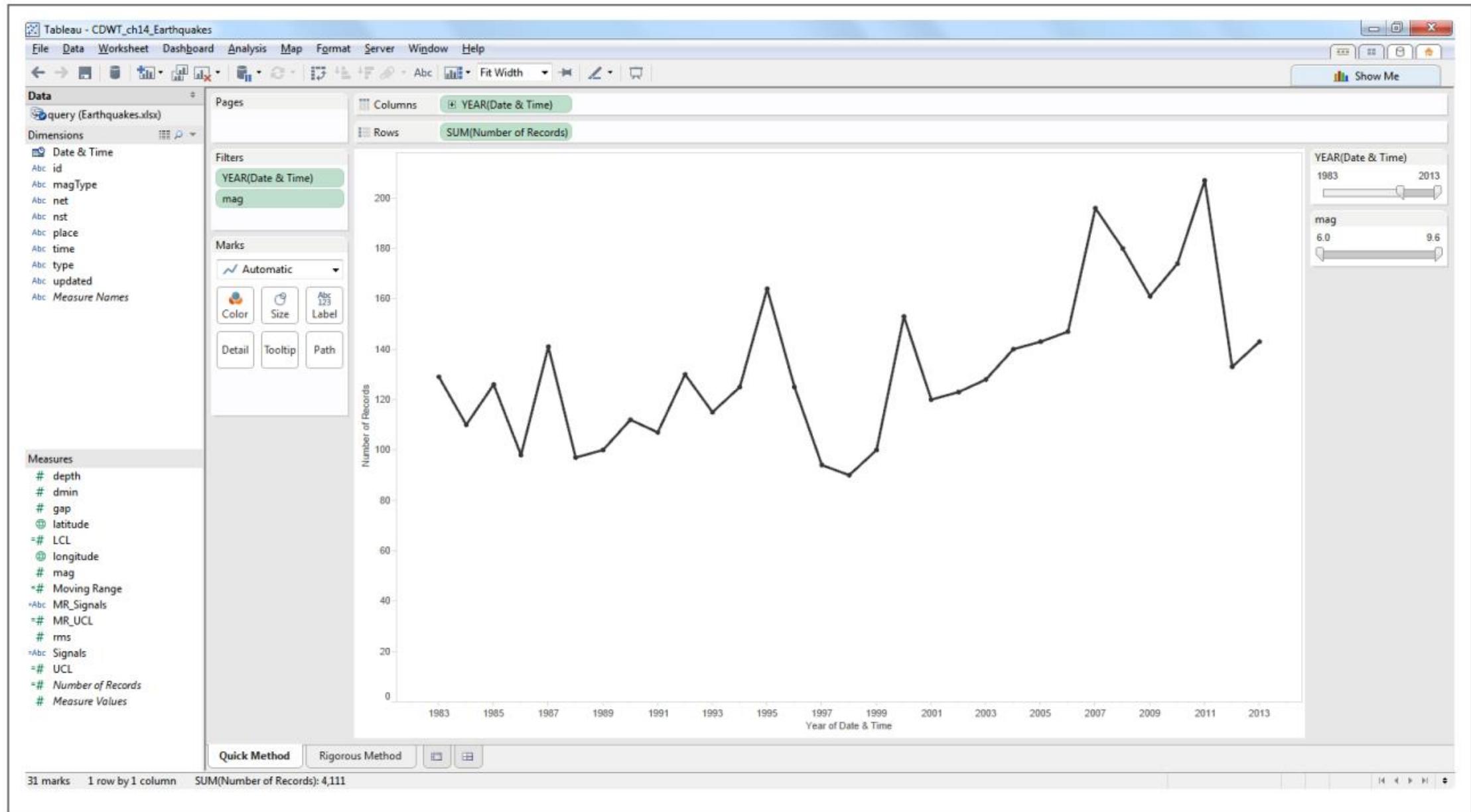


Figure 7-7. A simple timeline of the number of annual earthquakes

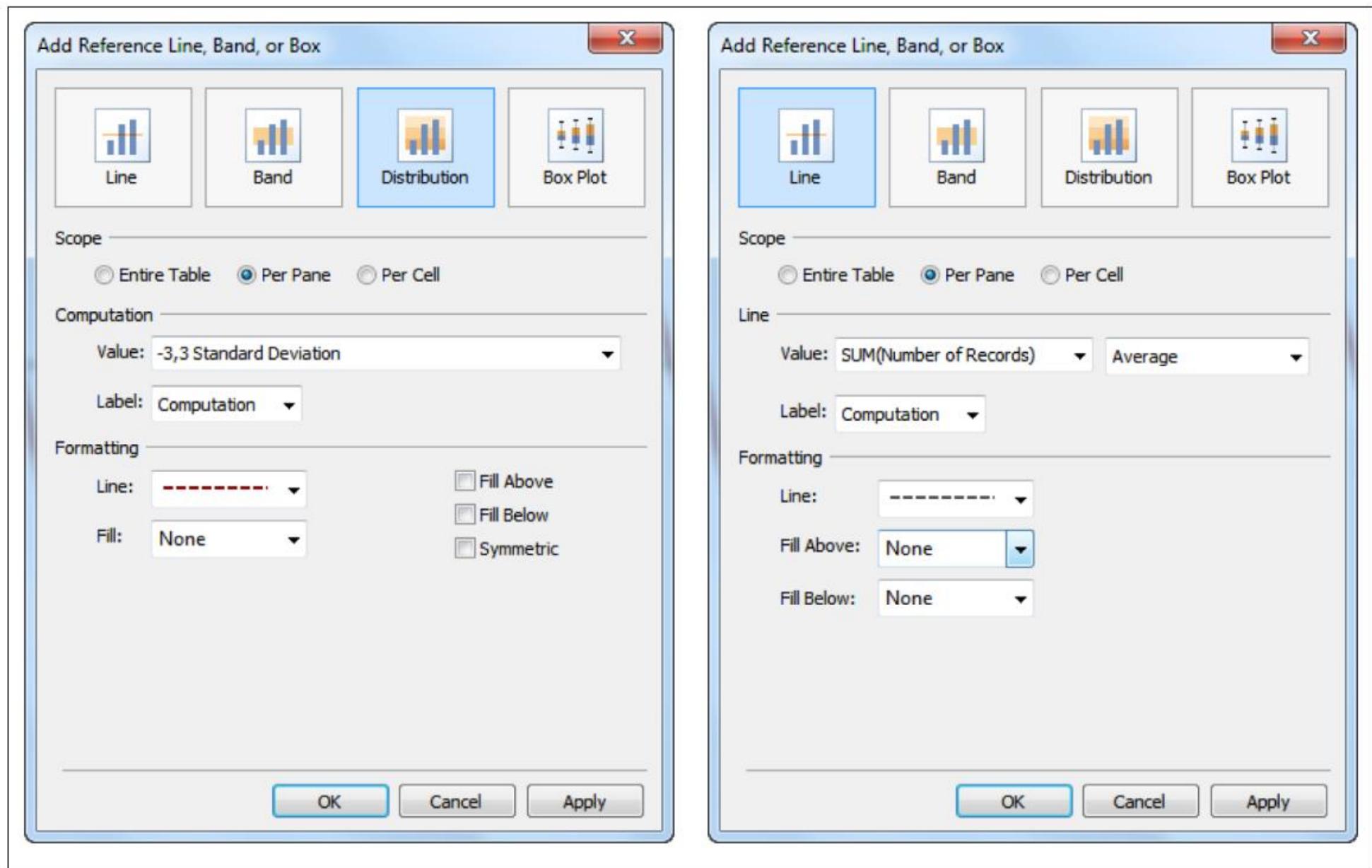
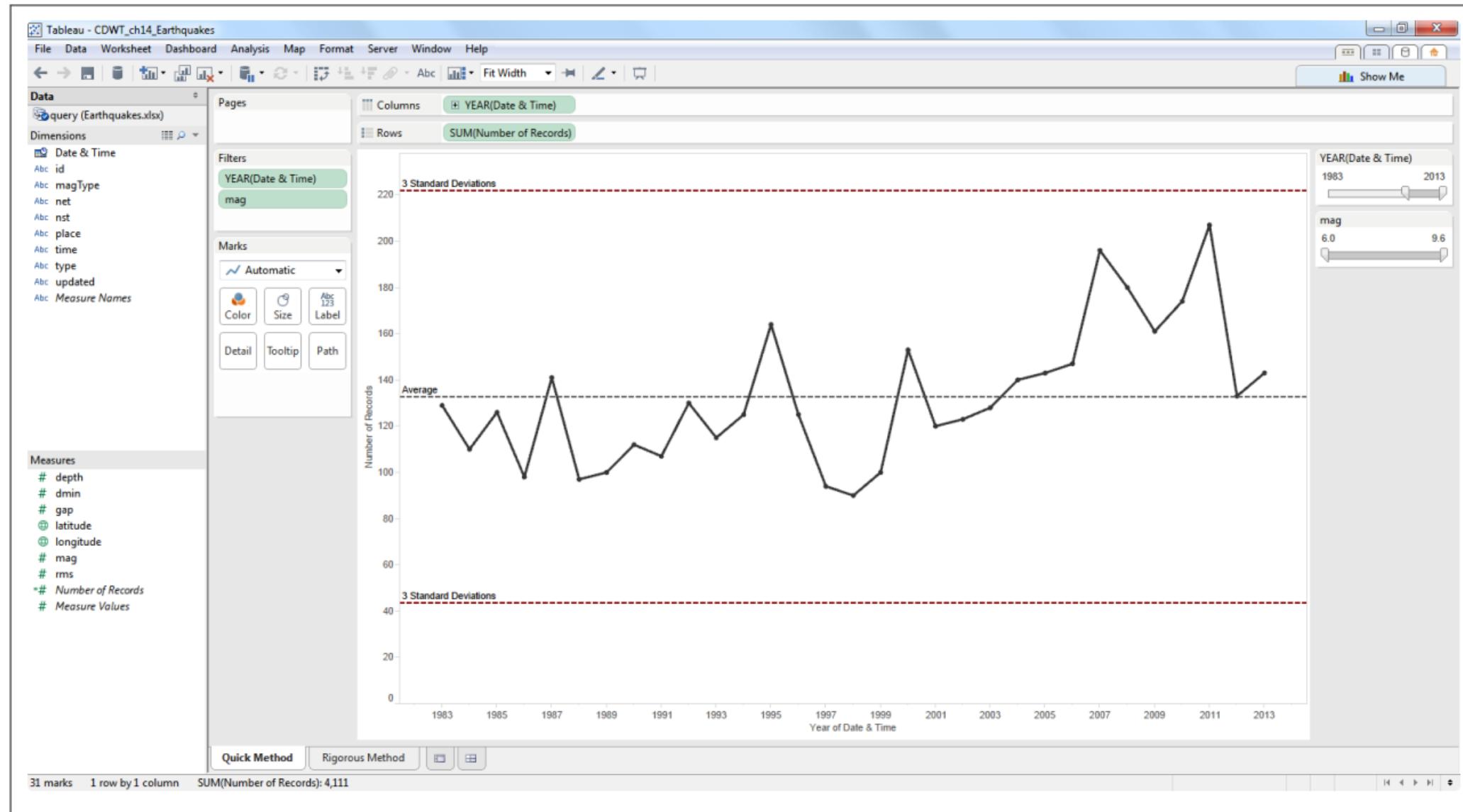


Figure 7-8. Adding reference lines to the line chart



*Figure 7-9. Simple control chart of annual earthquakes of magnitude 6.0 or greater*



# Variation over time-Control charts

- If we change from YEAR to MONTH, then the control chart changes to show several points above the 3-sigma line, including a sharp outlier in March 2011 corresponding to the Great East Japan earthquake, as shown in Figure 7~10
- Also note that the lower limit is not real. It's below 0, and it's not possible to have a negative number of earthquakes recorded

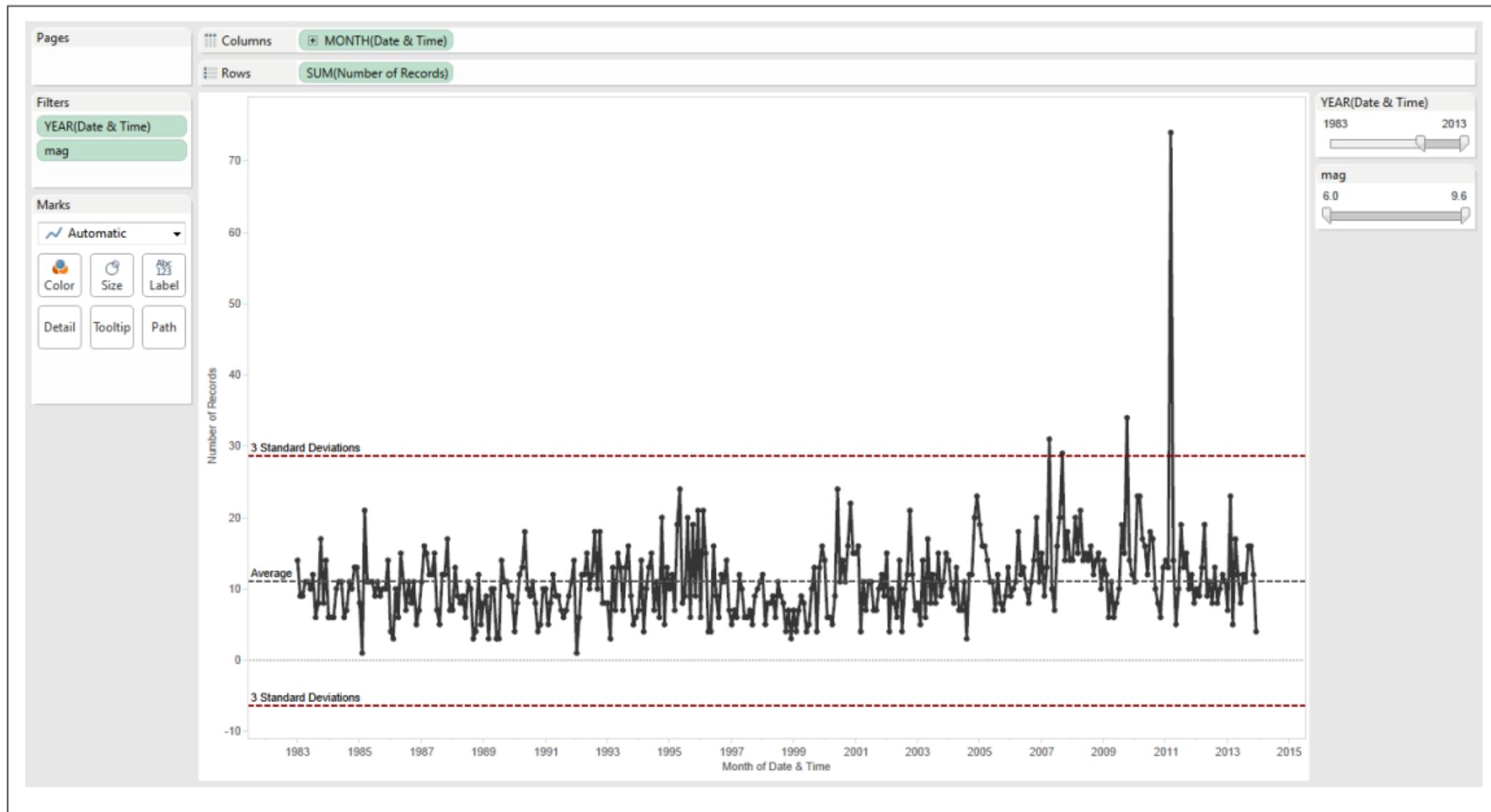


Figure 7-10. The simple control chart showing monthly counts of worldwide earthquakes

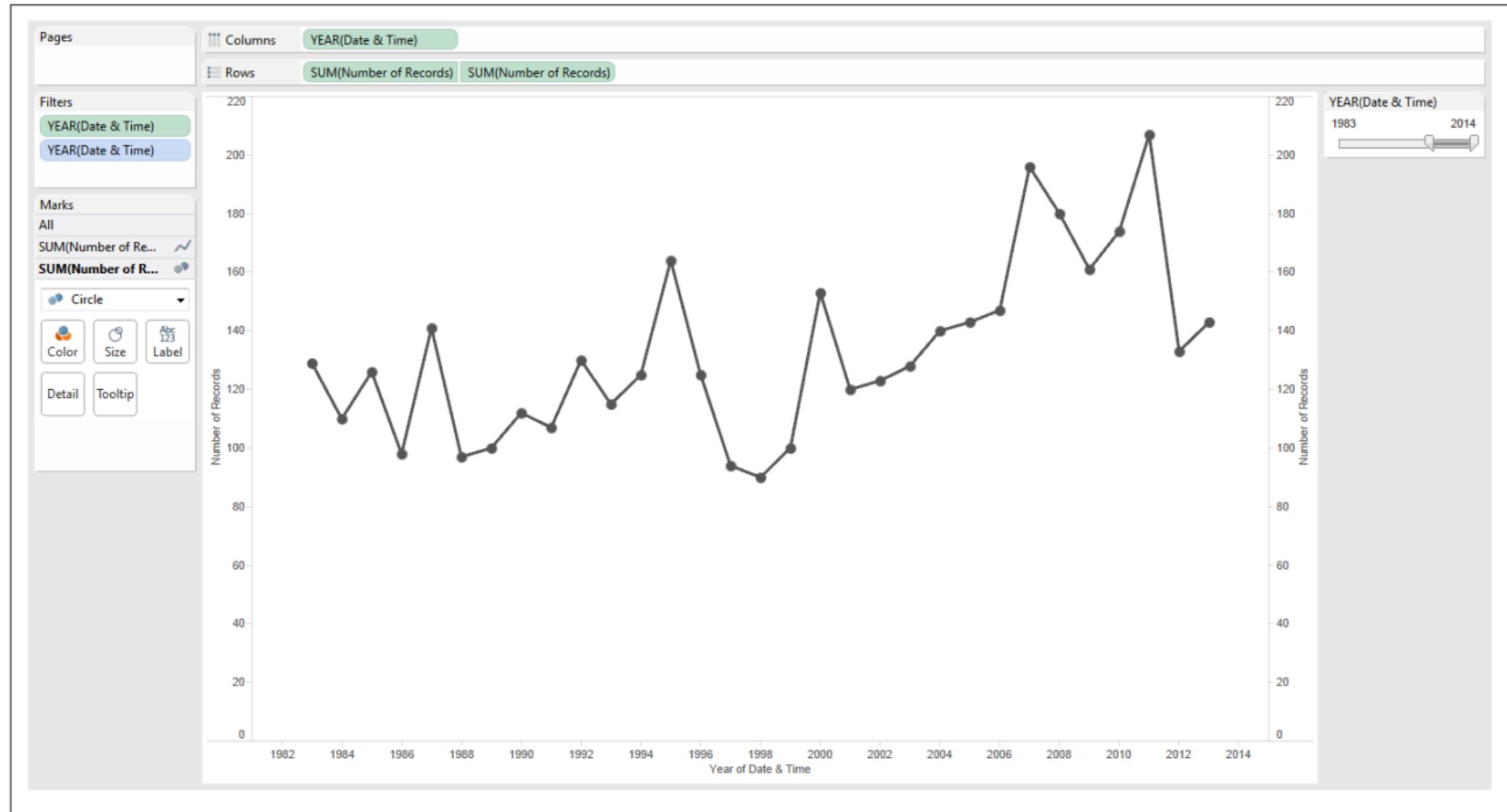


# Variation over time-Control charts

## The rigorous method:

- Create a new sheet and begin with Step 1 of the quick method outlined in the previous section to establish a basic timeline.
- Duplicate the SUM(Number of Records) and generate a dual-axis plot with synchronized axes.

- Represent the first set of marks as a line and the second set as circles, as illustrated in Figure 7~11.
- Additionally, introduce extra elements like a "Moving Range" timeline, which displays the absolute value of the change from one quake to another.



*Figure 7-11. Dual-axis timeline of annual earthquake count*

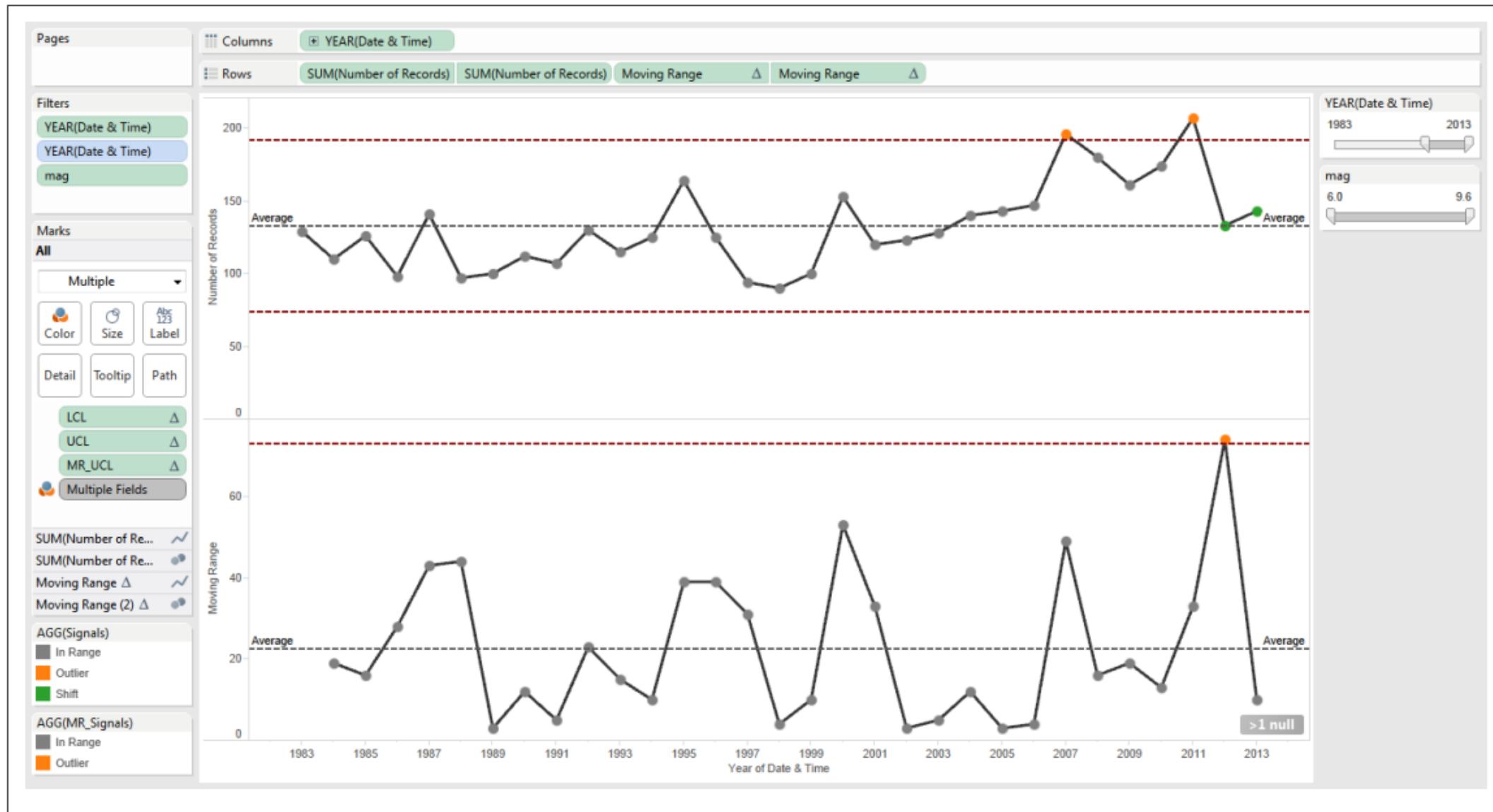


Figure 7-15. The rigorous control chart showing worldwide earthquake count by year



# Variation and Uncertainty

1

Respecting variation

2

Variation over time-Control charts

3

Understanding uncertainty



# Understanding uncertainty

- When we present data to an audience, it's important to make clear whether we are presenting data from the entire population, or from just a sample of the population.

- To illustrate the point, let's consider a fictional case of chess club participation rates of students in ten different cities in the peaceful country of Chesslandia.
- The President of Chesslandia, Garry Fischer, wanted to know whether students in his country were involved in chess club or not, and which cities were more successful in getting students to participate.

	A	B	C	D
1	Grade	Town	In Chess Club	Not in Chess Club
2	5th-6th grade	Kingston	29	40
3	5th-6th grade	Queensville	39	30
4	5th-6th grade	Rooktown	39	31
5	5th-6th grade	Bishop Village	18	32
6	5th-6th grade	Knightfield	35	33
7	5th-6th grade	Pawnford	36	18
8	5th-6th grade	Castleborough	38	22
9	5th-6th grade	Checkshire	33	31
10	5th-6th grade	Fianchettoberg	32	33
11	5th-6th grade	Gambitopolis	31	30
12	3rd-4th grade	Kingston	14	28
13	3rd-4th grade	Queensville	16	22
14	3rd-4th grade	Rooktown	13	21
15	3rd-4th grade	Bishop Village	7	26
16	3rd-4th grade	Knightfield	15	26
17	3rd-4th grade	Pawnford	14	21
18	3rd-4th grade	Castleborough	22	14
19	3rd-4th grade	Checkshire	15	27
20	3rd-4th grade	Fianchettoberg	14	27
21	3rd-4th grade	Gambitopolis	11	27
22	1st-2nd grade	Kingston	7	20
23	1st-2nd grade	Queensville	6	12
24	1st-2nd grade	Rooktown	8	20
25	1st-2nd grade	Bishop Village	4	17
26	1st-2nd grade	Knightfield	6	20
27	1st-2nd grade	Pawnford	6	12
28	1st-2nd grade	Castleborough	10	14
29	1st-2nd grade	Checkshire	7	20
30	1st-2nd grade	Fianchettoberg	8	15
31	1st-2nd grade	Gambitopolis	6	16

Figure 7-17. Spreadsheet of results of chess club student survey

# Chess Club Popularity in Chesslandia

Bar lengths proportional to % in chess club, bar widths proportional to the number of students surveyed. At first blush, it seems that we have a fairly large difference in chess club participation between students in different towns of Chesslandia.

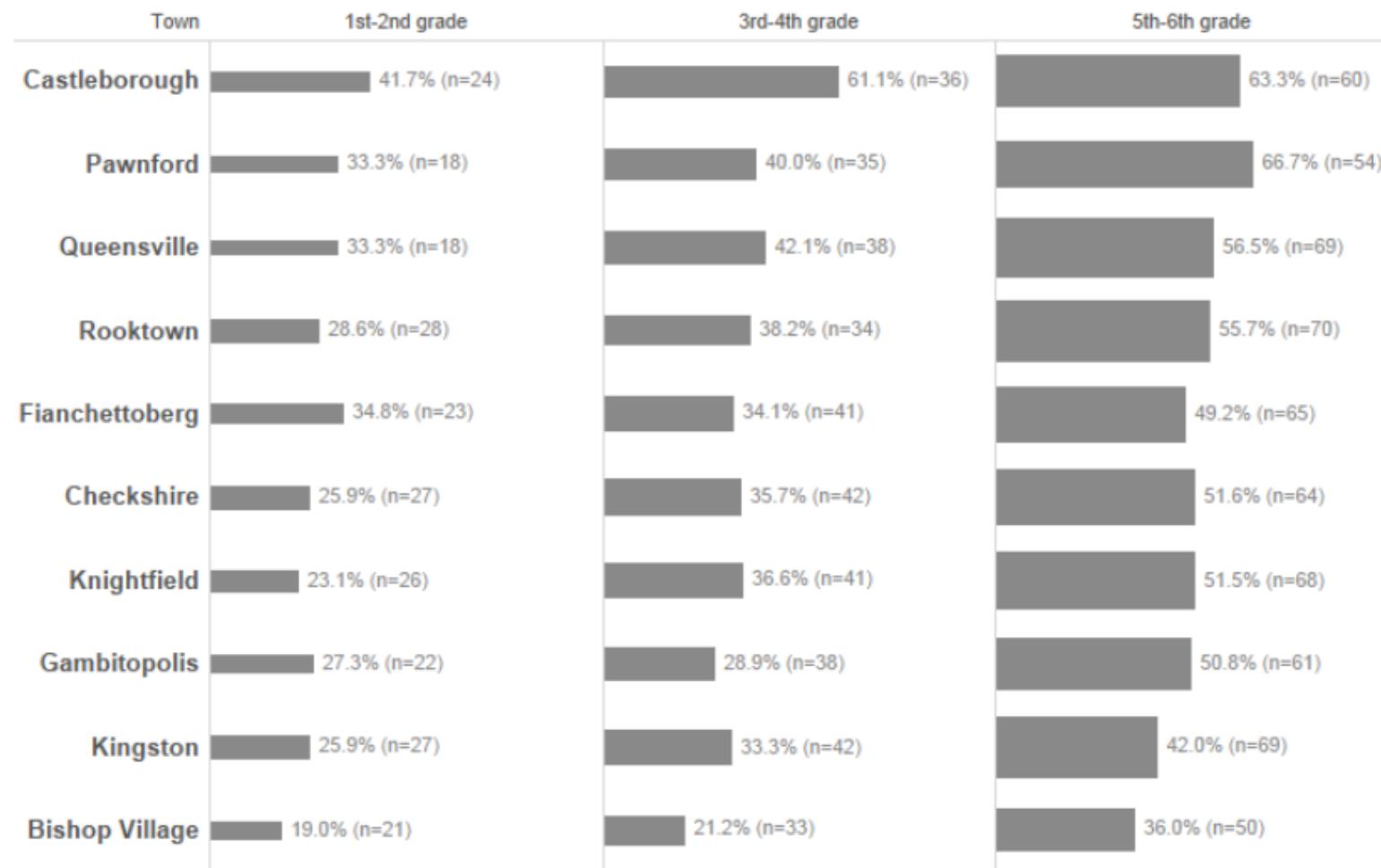


Figure 7-18. Visualization of the survey results

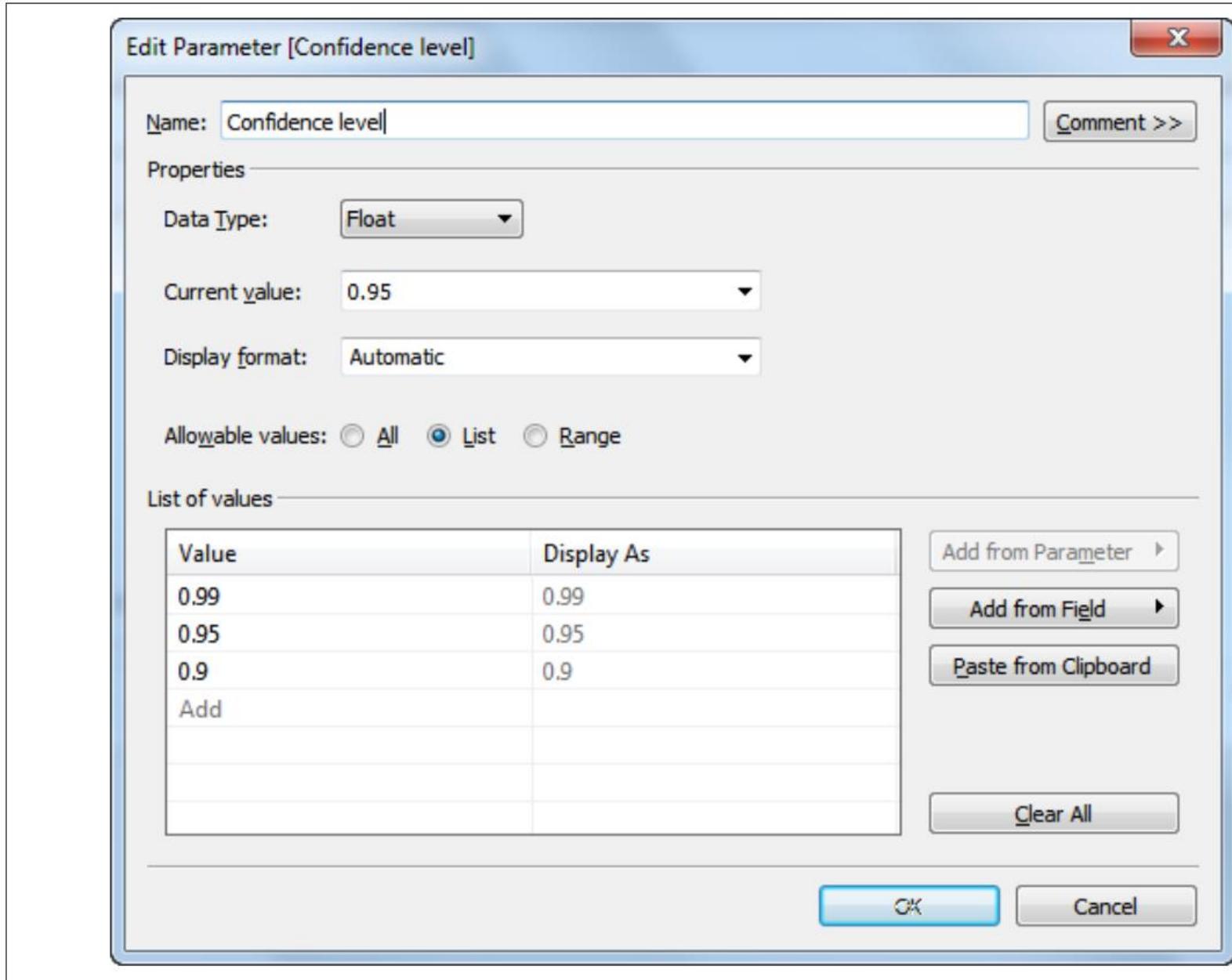


Figure 7-19. A new parameter to select desired confidence level



# Understanding uncertainty

Next, he needed to create several calculated fields to generate the error bars:

- % in Chess Club,  $p = [\text{In Chess Club}] / ([\text{In Chess Club}] + [\text{Not in Chess Club}])$
- Sample Size,  $n = [\text{In Chess Club}] + [\text{Not in Chess Club}]$
- Standard Error =  $\text{SQRT}(([\% \text{ in Chess Club}] * (1 - [\% \text{ in Chess Club}])) / [n])$
- $z_{\text{upper}} = \text{CASE } [\text{Confidence level}]$ 
  - — WHEN 0.99 THEN 2.575829
  - — WHEN 0.95 THEN 1.959964
  - — WHEN 0.90 THEN 1.644854



# Understanding uncertainty

- Margin of Error = [Standard Error]\*[z upper]
- Lower limit = [% in Chess Club]–[Margin of Error]
- Upper limit = [% in Chess Club]+[Margin of Error]
- Error Bar Line = [Upper limit]–[Lower limit]
- $np = [n]*[% \text{ in Chess Club}]$

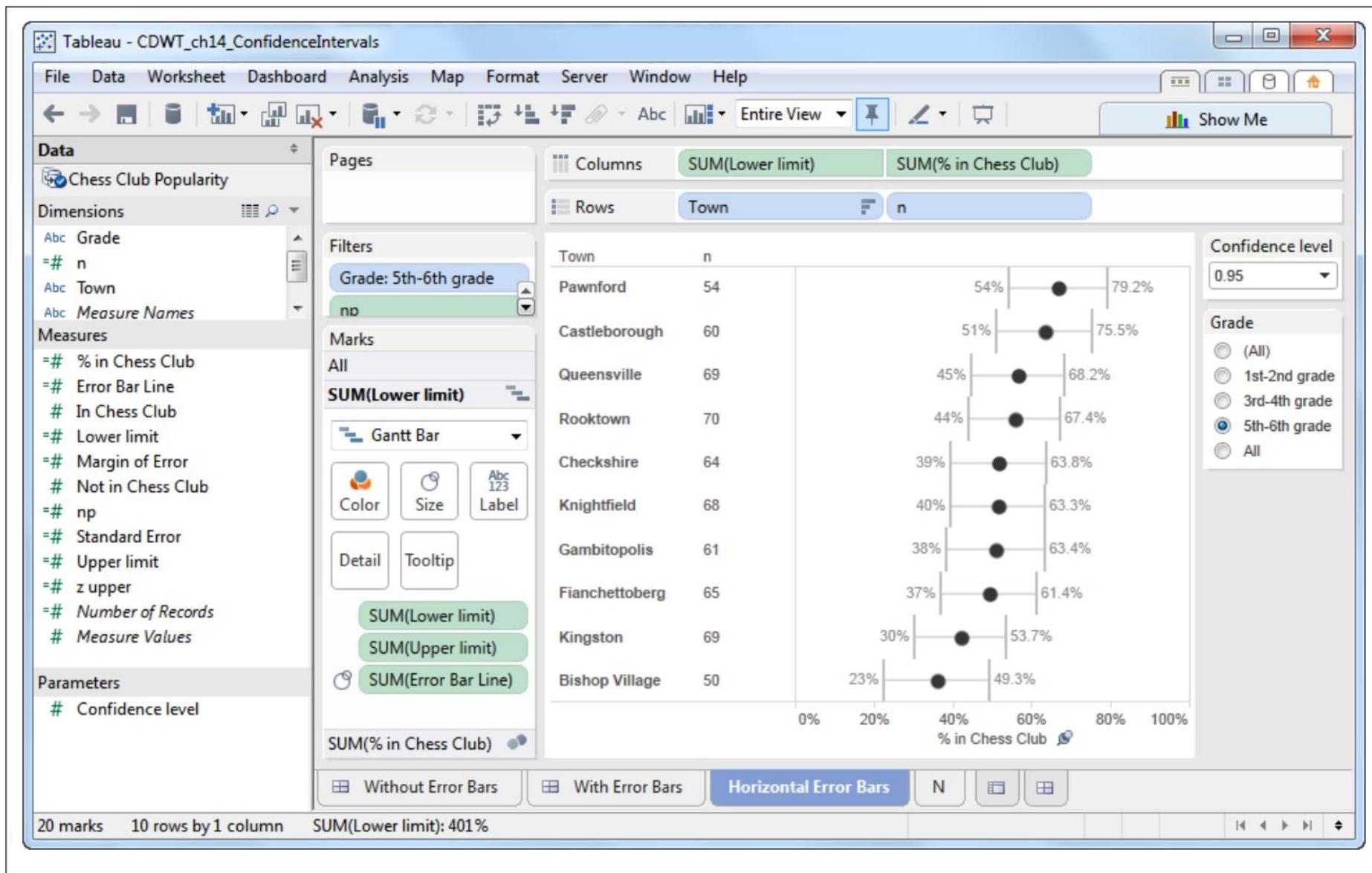


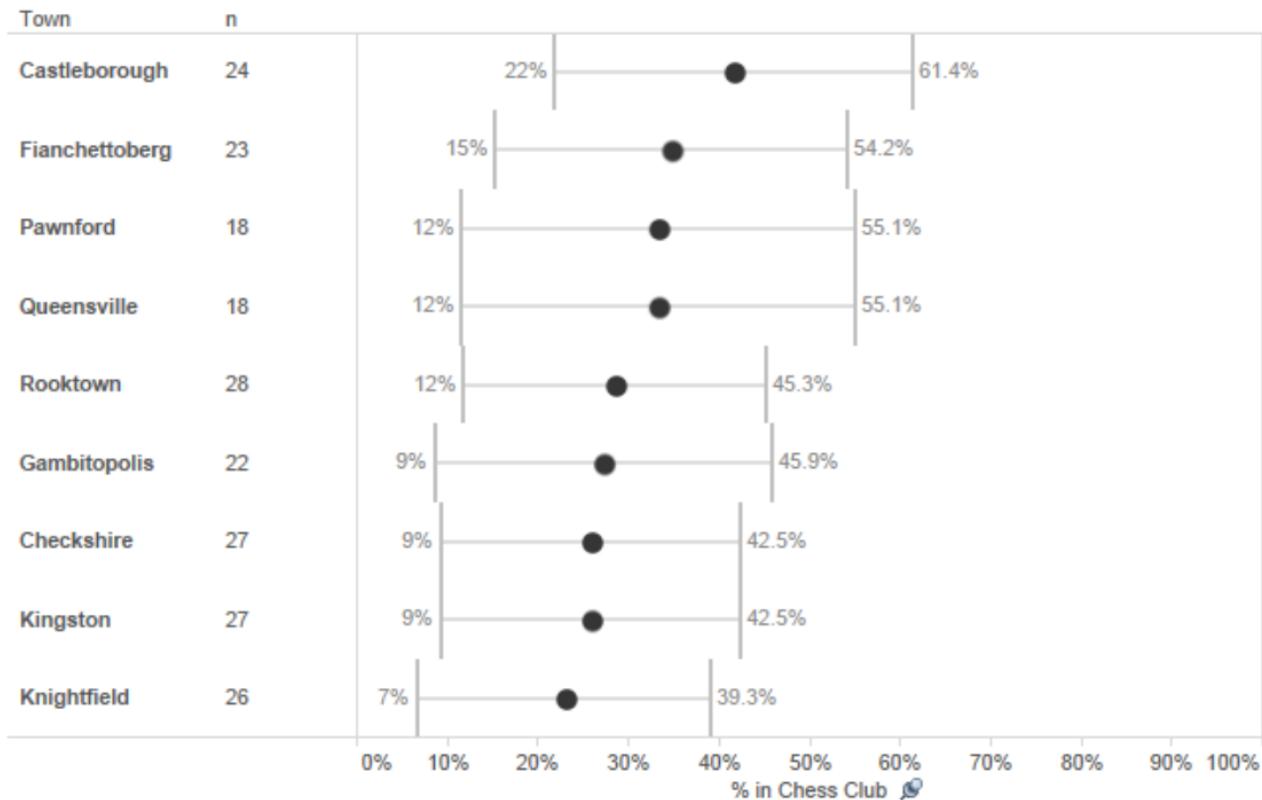
Figure 7-20. A binomial probability distribution for the chess club survey results

# Chess Club Popularity in Chesslandia

When confidence intervals are taken into consideration, it becomes clear that we can say very little with confidence about which city is higher and which is lower, even within grade levels.

Confidence level  
0.95 ▾

Grade  
1st-2nd grade ▾



Data Source | Fictional

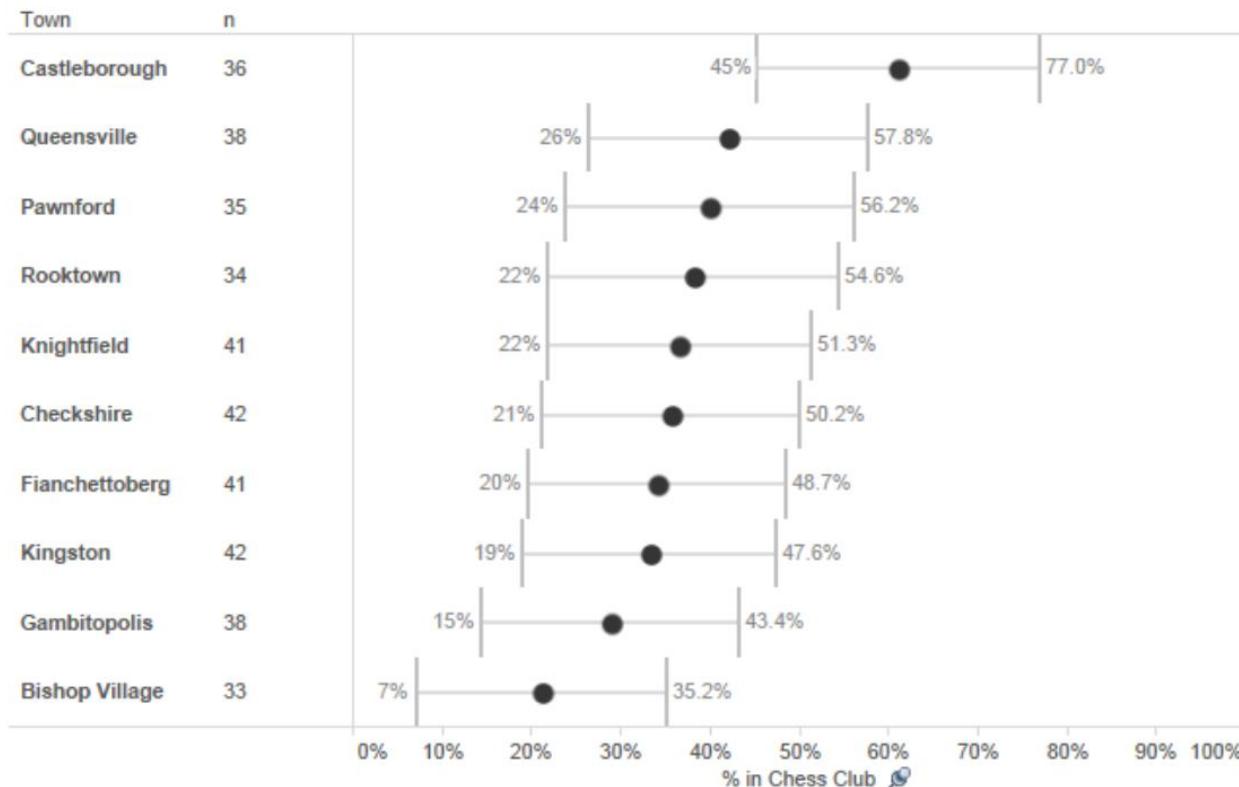
Mar 2014

Figure 7-21. 95% confidence intervals for first- and second-grade chess club participation

# Chess Club Popularity in Chesslandia

When confidence intervals are taken into consideration, it becomes clear that we can say very little with confidence about which city is higher and which is lower, even within grade levels.

Confidence level  
0.95 ▾  
Grade  
3rd-4th grade ▾



Data Source | Fictional

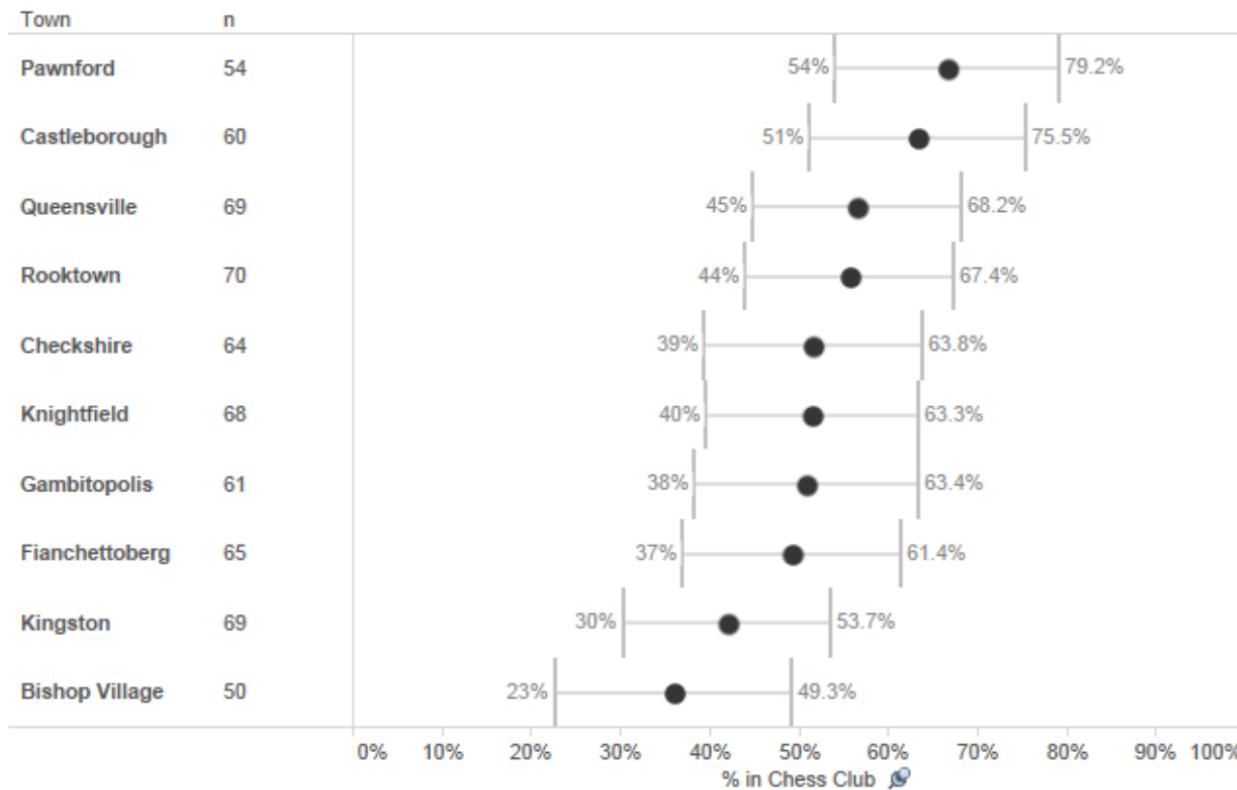
Mar 2014

Figure 7-22. 95% confidence intervals for third- and fourth-grade chess club participation

# Chess Club Popularity in Chesslandia

When confidence intervals are taken into consideration, it becomes clear that we can say very little with confidence about which city is higher and which is lower, even within grade levels.

Confidence level  
0.95 ▾  
Grade  
5th-6th grade ▾



Data Source | Fictional

Mar 2014

Figure 7-23. 95% confidence intervals for fifth- and sixth-grade chess club participation



# DATA VISUALIZATION



# **UNIT - V (PART - I)**



# Proportions and Percentages

1

Scatterplots

2

Stacked Bars

3

Regression and Trend Lines

4

The Quadrant Chart



# Scatterplots

- It's great to hear that you appreciate scatterplots! They are indeed a powerful tool for visualizing and analyzing data, providing a quick and intuitive way to identify patterns, trends, and outliers.
- They create a two-dimensional plane in which a whole host of comparisons can be made in an instant.

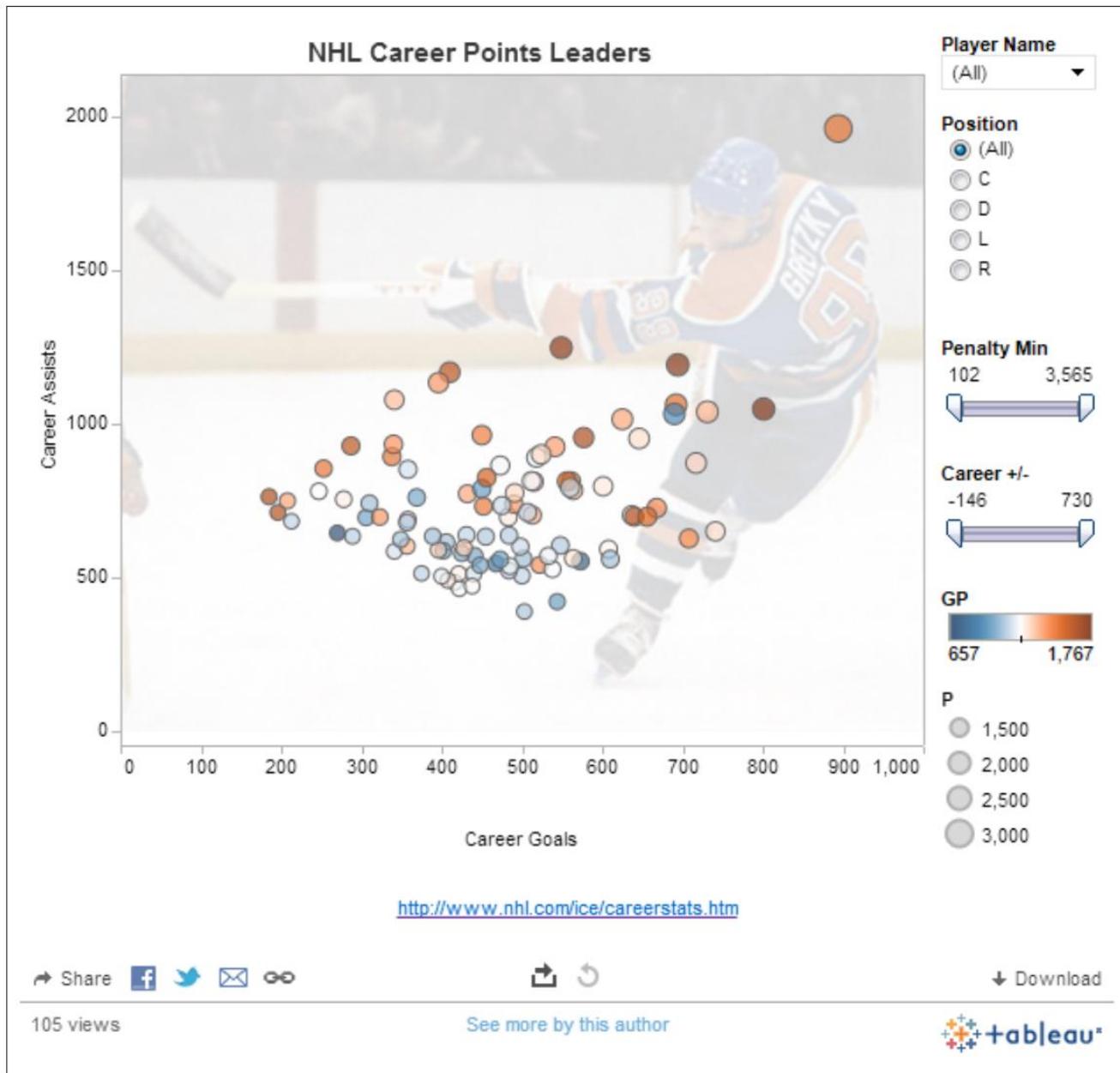


Figure 8-1. *My first Tableau Public visualization: a scatterplot dashboard of career stats (photo by B. Bennett/Getty Images)*



# Scatterplots

- Let's explore how to create this scatterplot.
- First, we'll connect to the spreadsheet that contains the top 100 players, which you can find online here.
- Once we've connected to the spreadsheet, it's a simple matter of Ctrl-selecting Player, G (for goals), and A (for assists), and then clicking on scatterplots in the Show Me panel, as shown in Figure 8-2

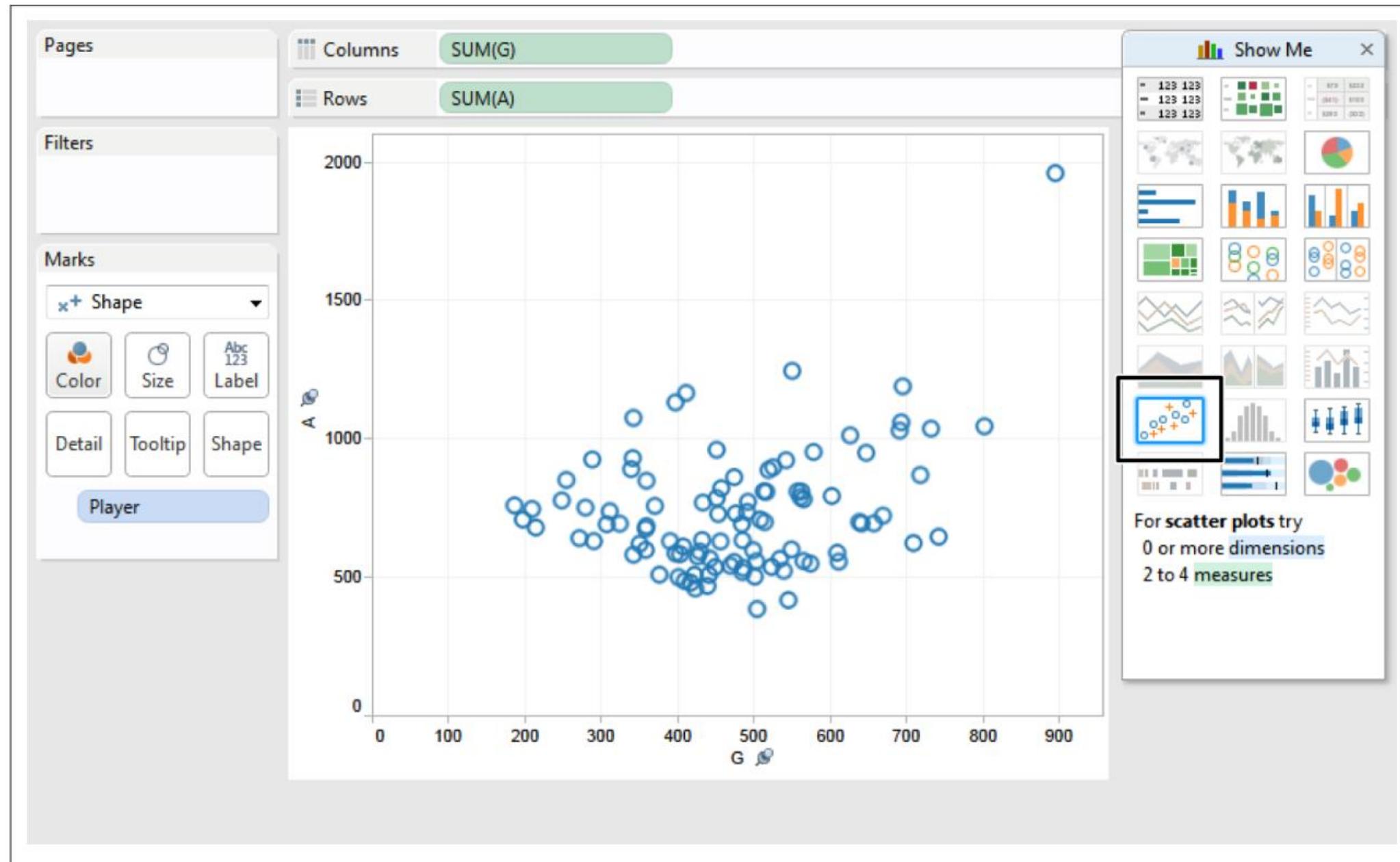


Figure 8-2. Creating a scatterplot using Show Me

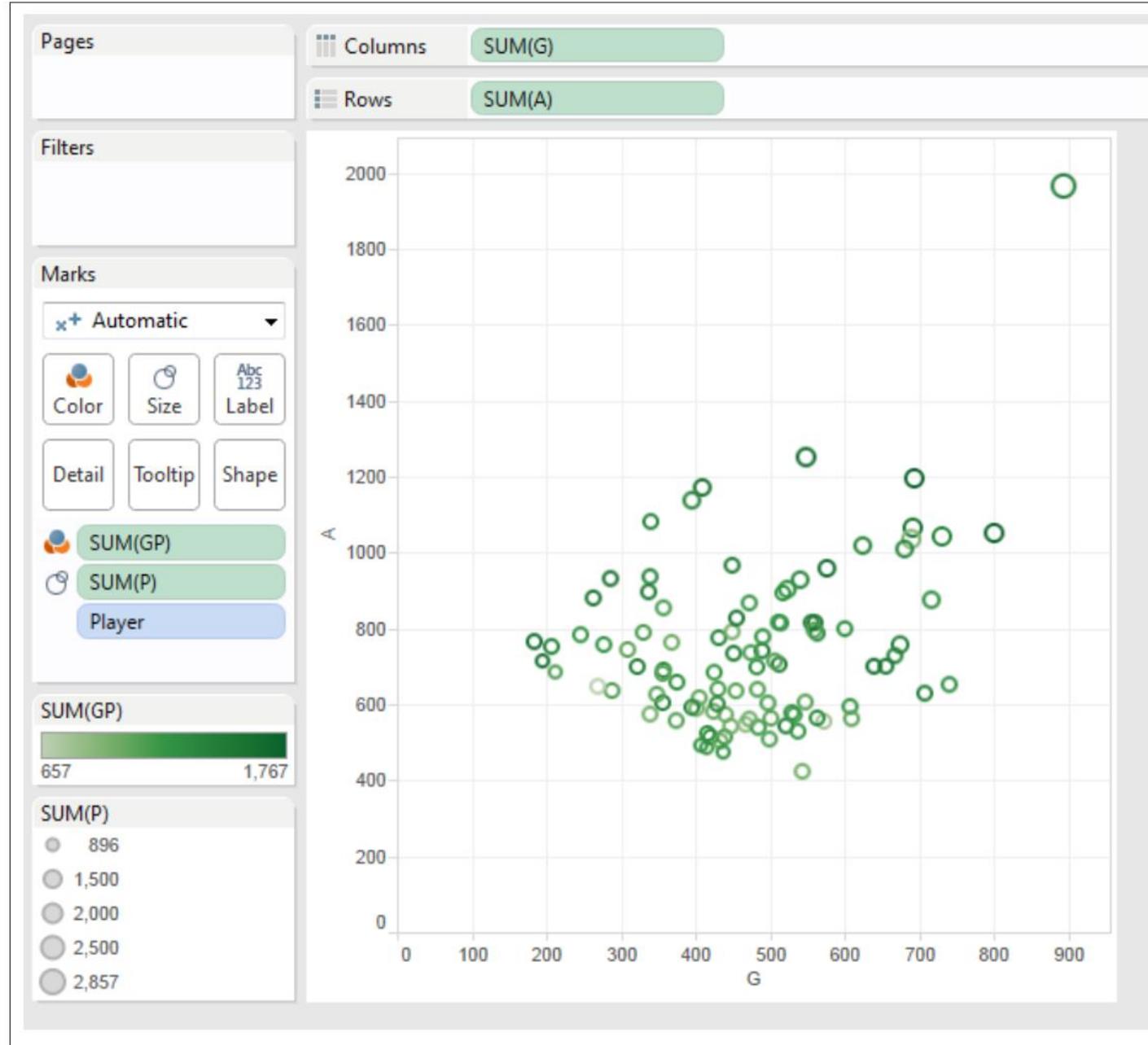


Figure 8-3. Scatterplot with added encodings for size and color



Figure 8-4. Additional formatting to the scatterplot



# Scatterplots

## Who Is Who?

- There are three ways to communicate who is who: labels, tooltips, and annotations.
- Let's consider them one at a time

### Labels:

- If we take Player from the Dimensions panel and drag it onto the Label shelf, Tableau attempts to add as many labels as it can without creating a messy view.
- The result is shown in Figure 8-5



Figure 8-5. Adding labels to scatterplots



# Scatterplots

## Tooltips:

### Who Is Who?

- There are three ways to communicate who is who: labels, tooltips, and annotations.
- Let's consider them one at a time

- What about the other points without labels? How can we see who they are?
- Tableau has a great feature called Tooltips, which appear when a person interacting with the chart hovers the mouse cursor over an individual mark (or circle, in this case).

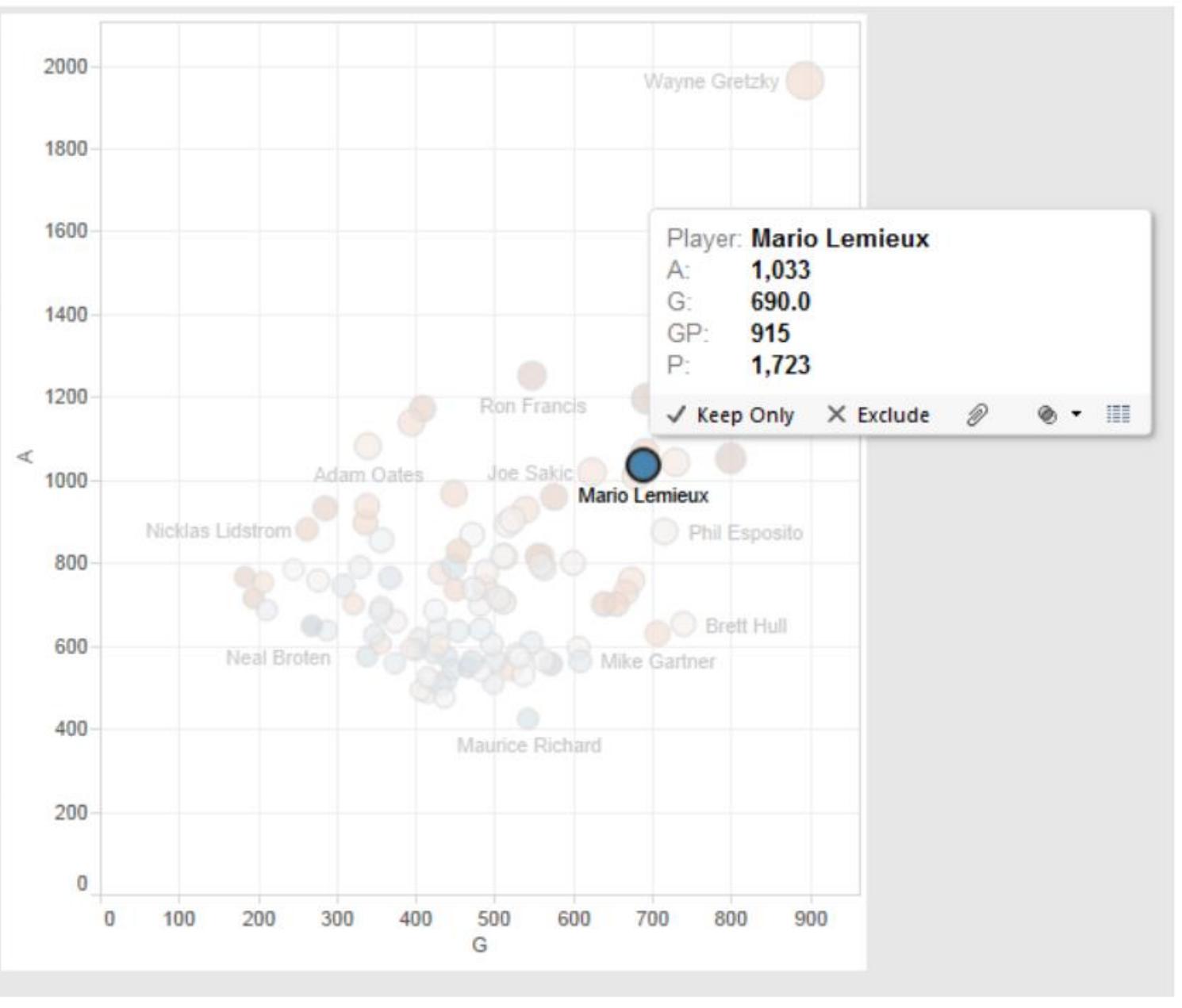


Figure 8-6. Hovering and clicking reveals tooltips

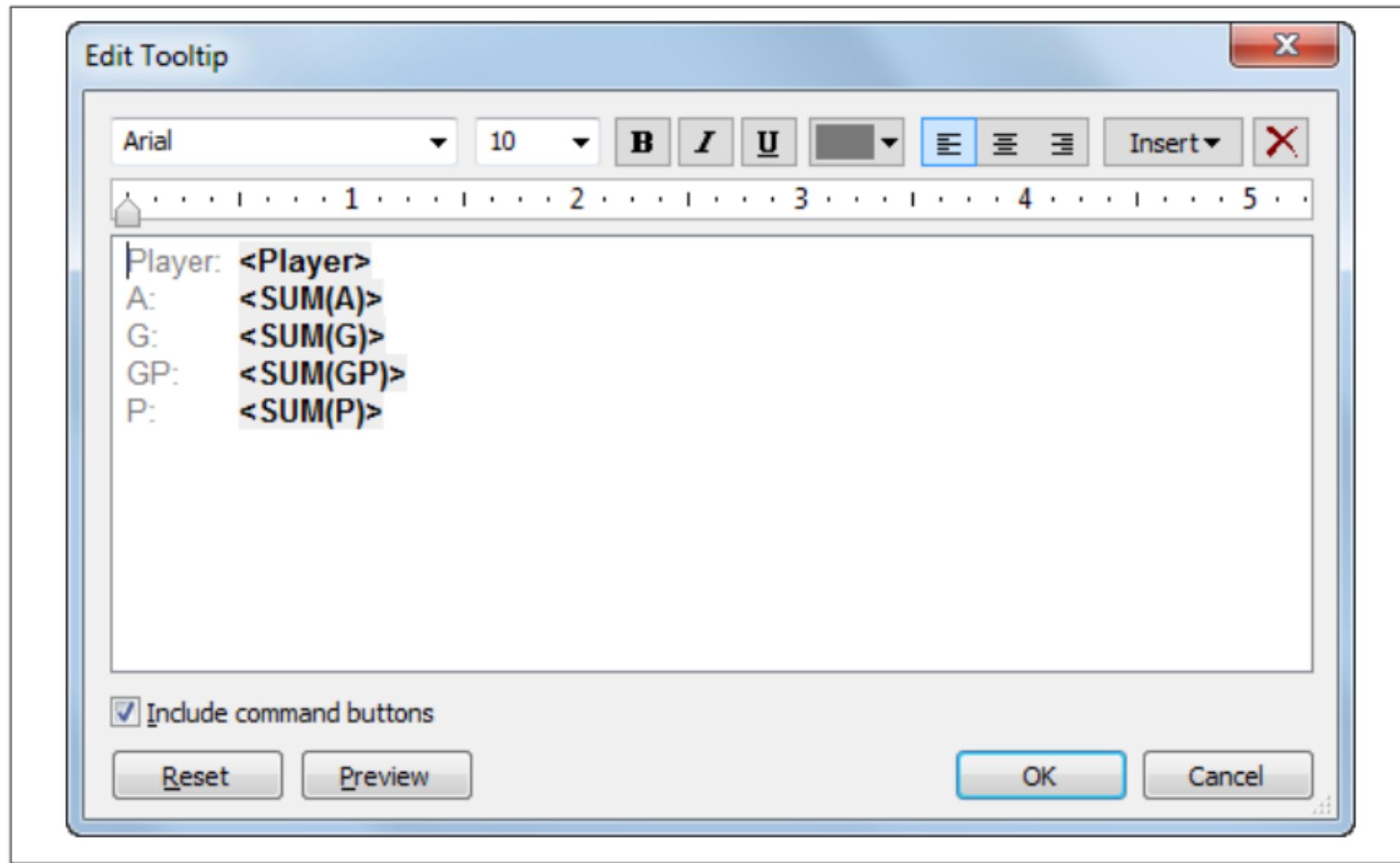


Figure 8-7. Editing tooltips



# Scatterplots

## Who Is Who?

- There are three ways to communicate who is who: labels, tooltips, and annotations.
- Let's consider them one at a time

## Annotations:

- Remove Player Labels
- Annotate Specific Data Points
- Repeat for Mario Lemieux:

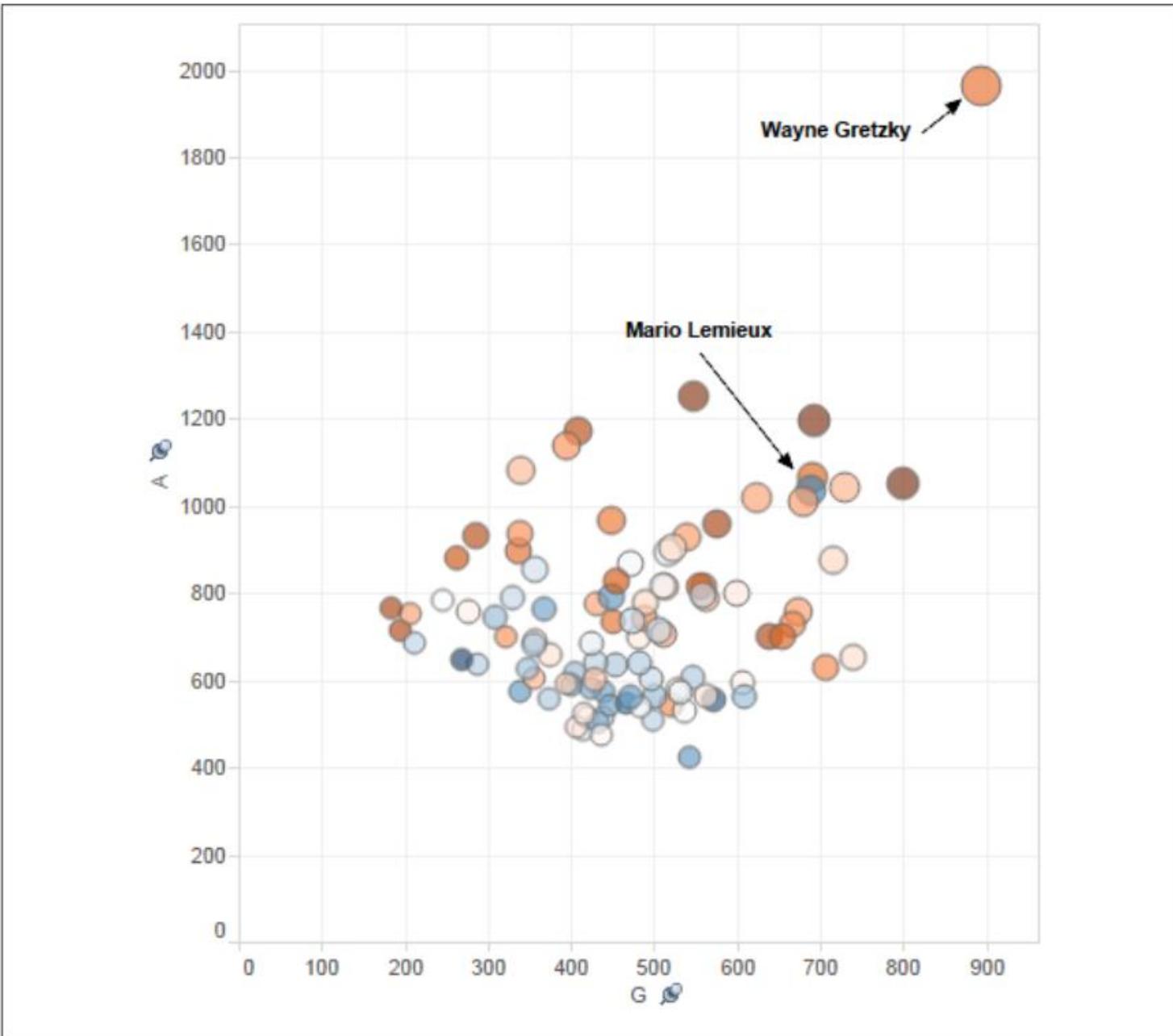


Figure 8-8. Adding annotations to scatterplots



# Scatterplots

## Making it Exploratory

- To add the three Quick Filters to the view, right-click on the three fields (Pos, +/-, and PIM) one at a time, and select Show Quick Filter for each one.

- Tableau adds a Multiple Values (combo box) list for Pos and Range of Values sliders for +/- and PIM, as shown in Figure 8-9.

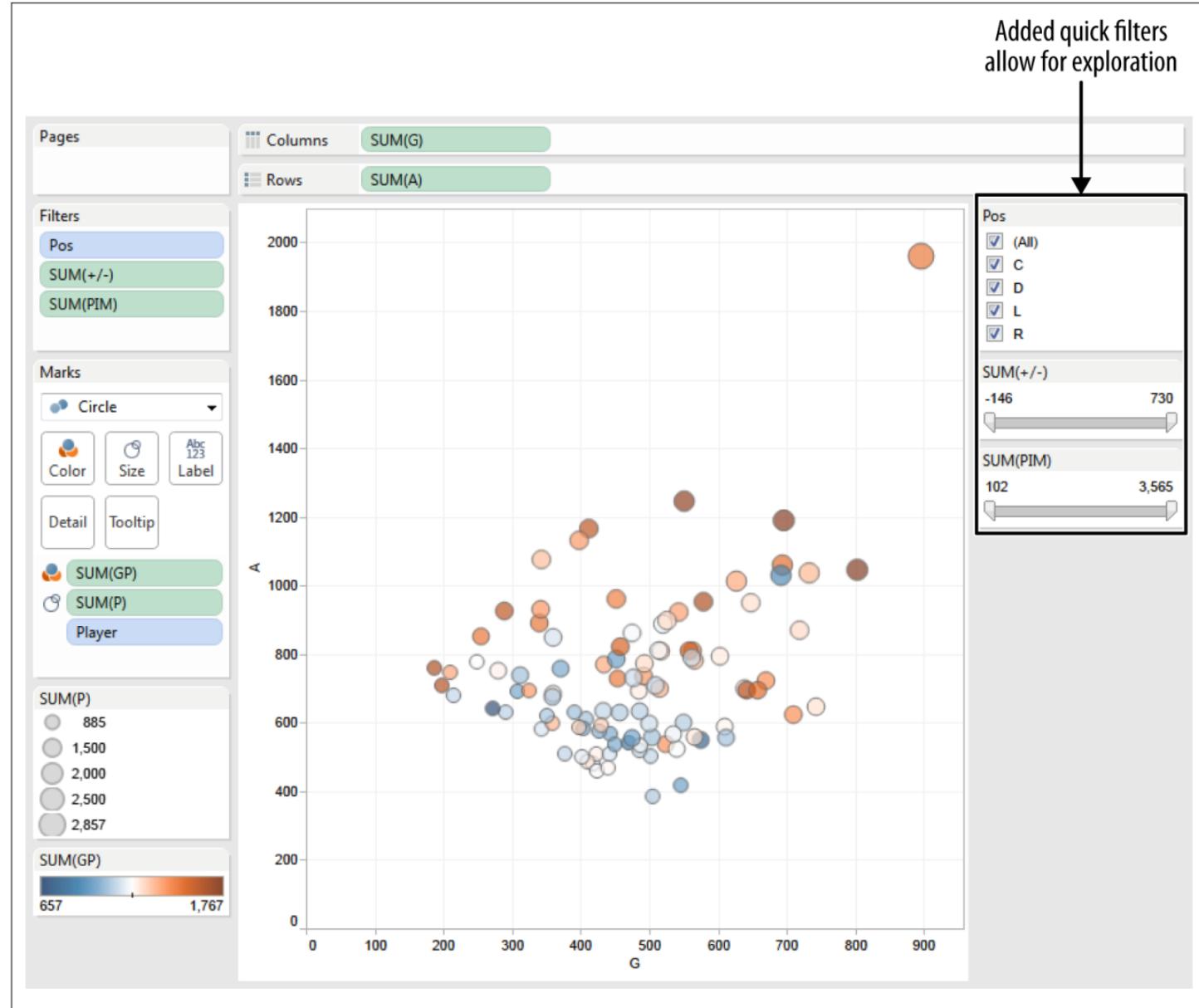


Figure 8-9. Quick Filters turn a scatterplot into an exploratory interactive



# Scatterplots

## Adding Background Images

- To add the three Quick Filters to the view, right-click on the three fields (Pos, +/-, and PIM) one at a time, and select Show Quick Filter for each one.

- Tableau adds a Multiple Values (combo box) list for Pos and Range of Values sliders for +/- and PIM, as shown in Figure 8-9.

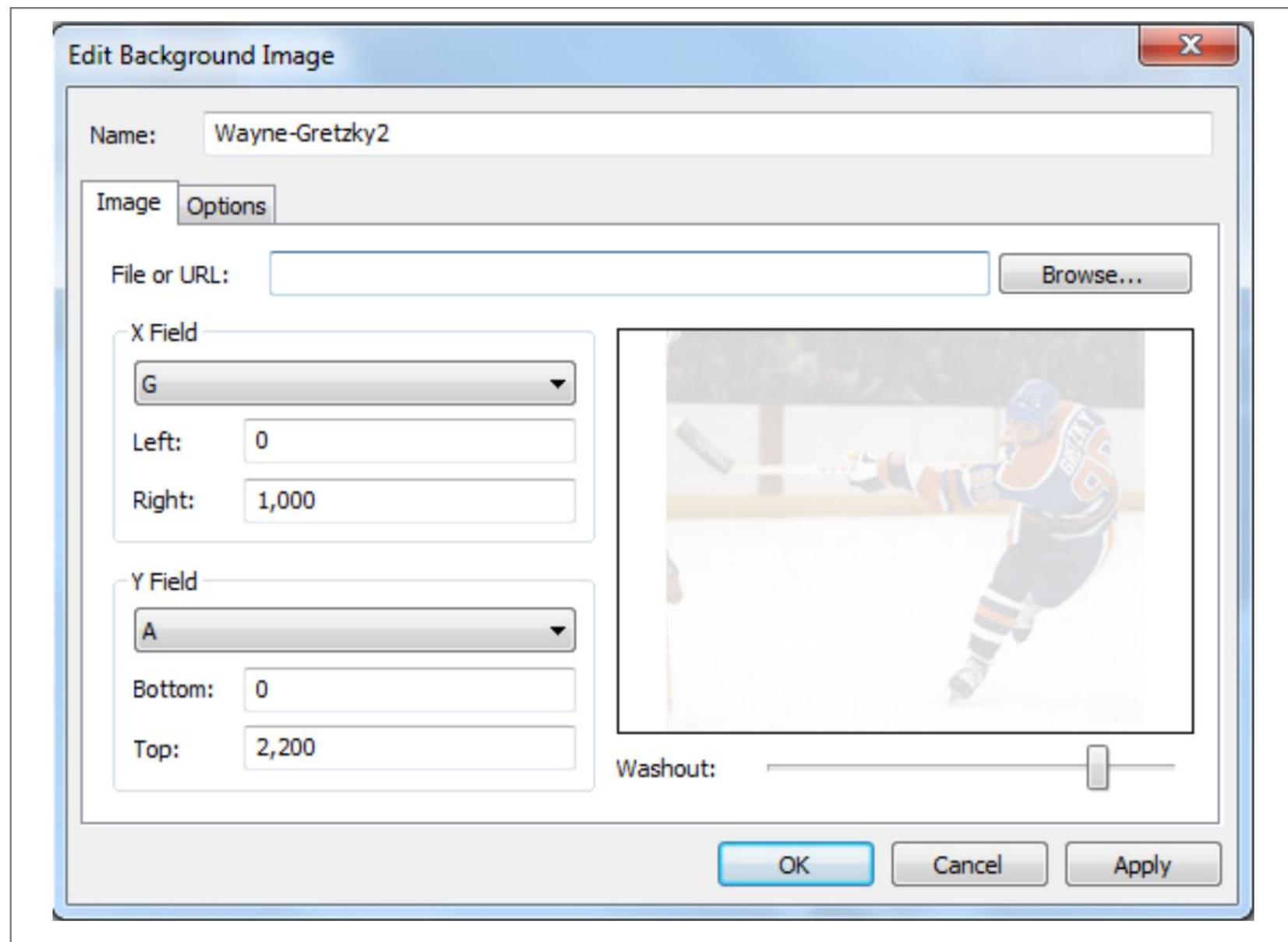


Figure 8-11. Adding a background image (photo by B. Bennett/Getty Images)



# Proportions and Percentages

1

Scatterplots

2

Stacked Bars

3

Regression and Trend Lines

4

The Quadrant Chart



# Stacked Bars

- Scatterplots aren't the only way to show multiple quantities in the same view. Another visualization type we can use is the stacked bar chart.
- Let's create a stacked bar chart as we explore a different angle of the data: per game rates

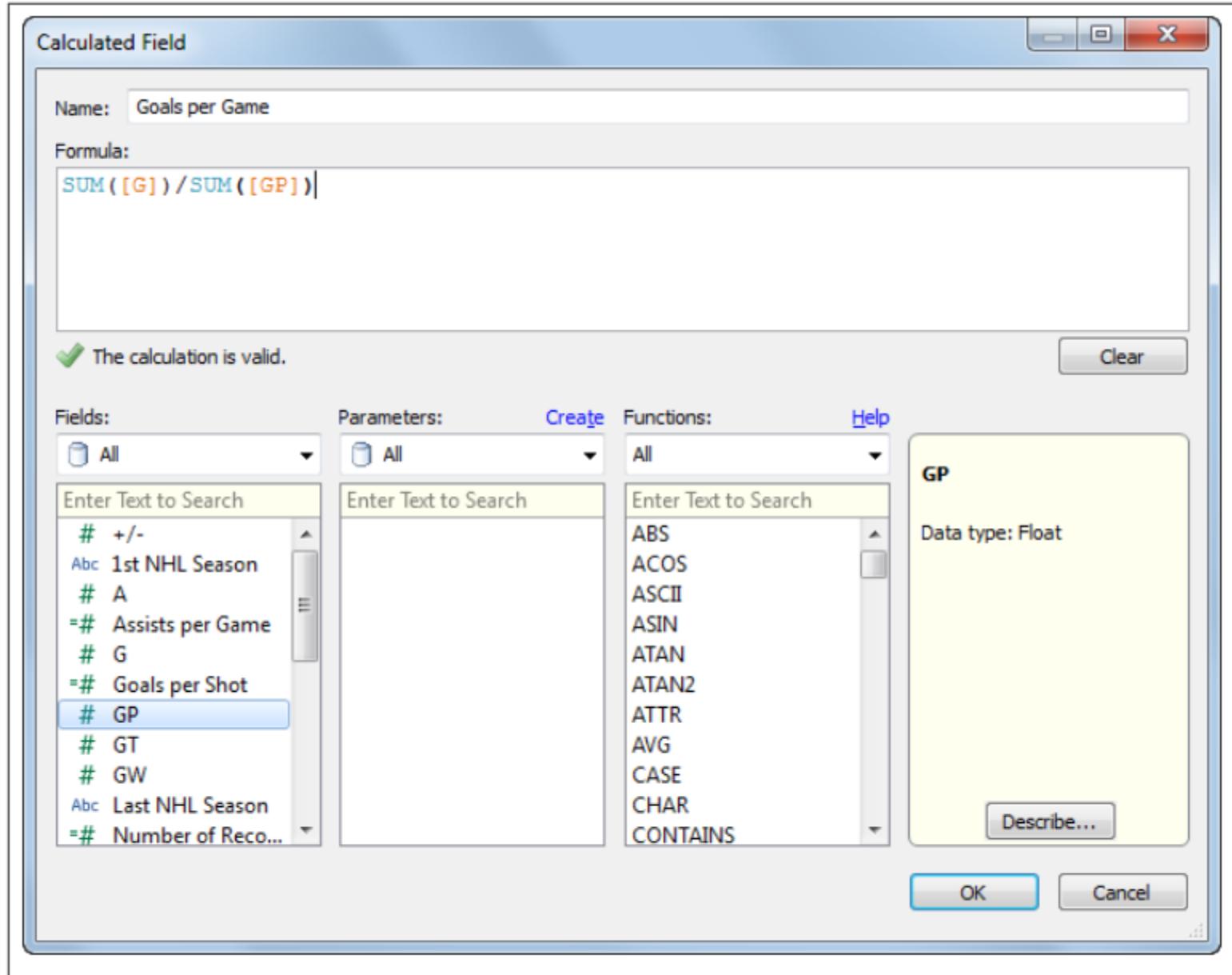


Figure 8-12. Creating per game rates with Calculated Fields

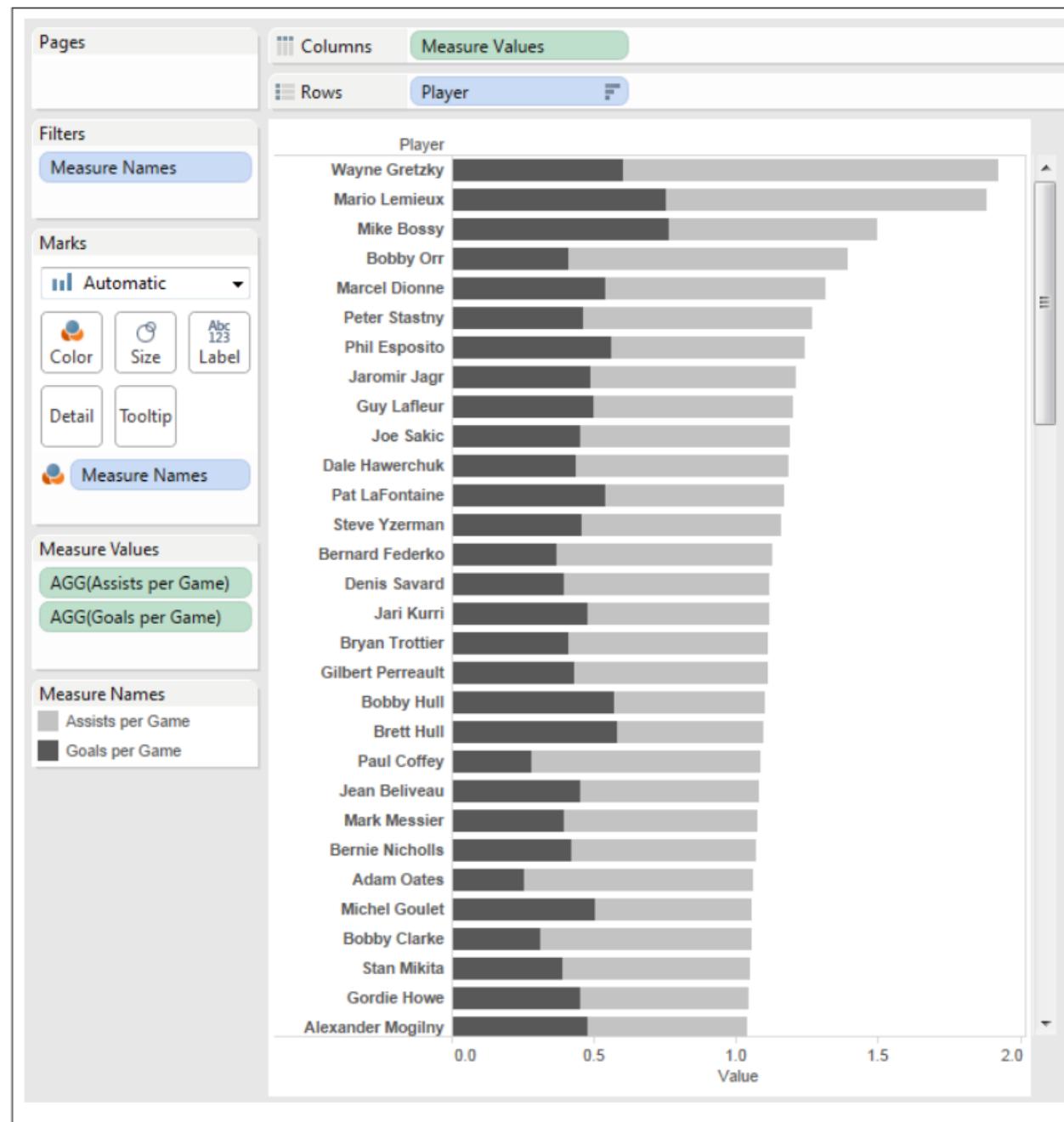


Figure 8-13. Creating a stacked bar from multiple Measures



# Proportions and Percentages

1

Scatterplots

2

Stacked Bars

3

Regression and Trend Lines

4

The Quadrant Chart



# Regression and Trend Lines

- Regression analysis can be visualized by adding a trend line to a scatterplot, and the indication of how well the points fit the trend line is known as the coefficient of determination.

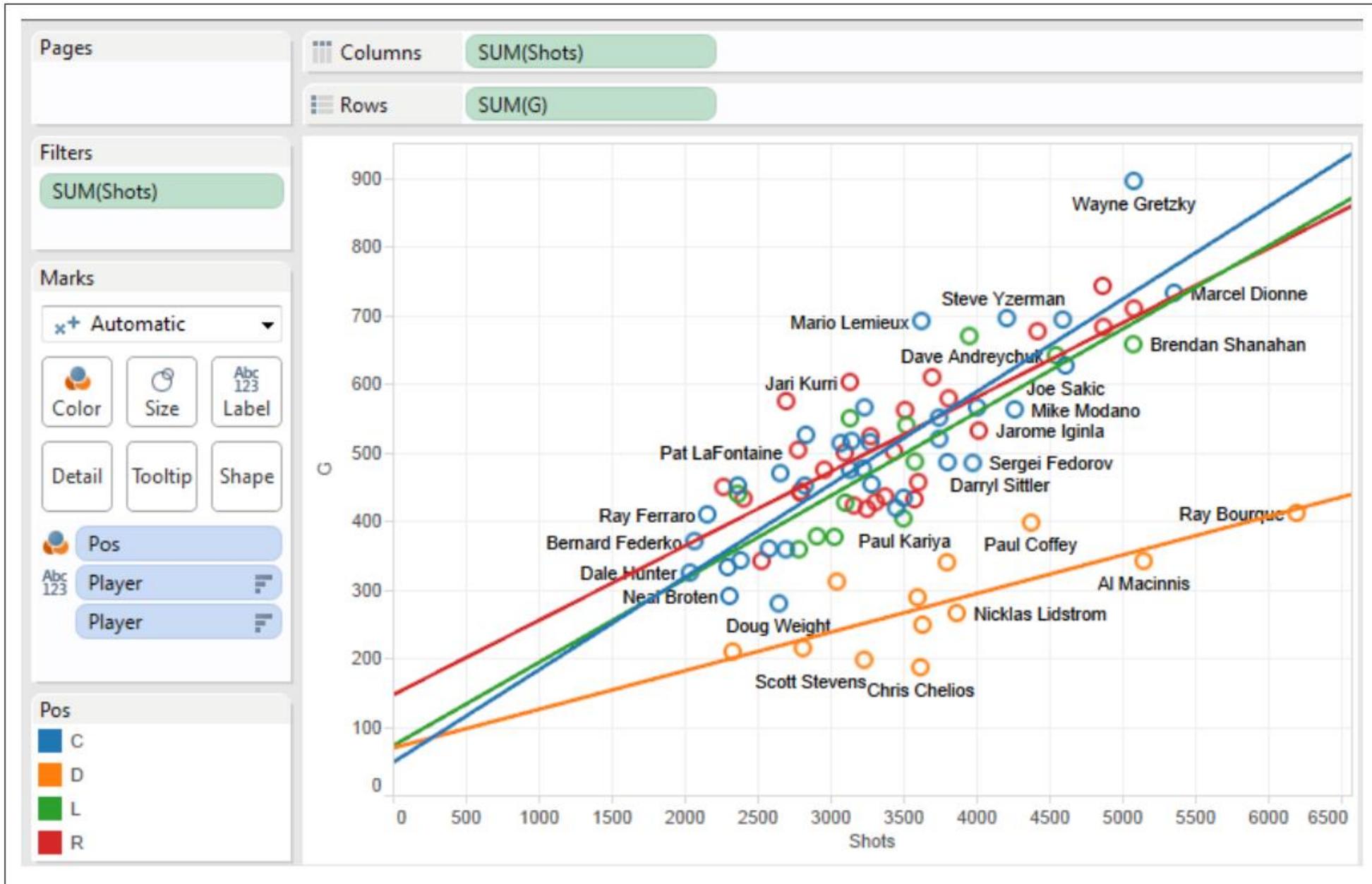


Figure 8-16. Scatterplot with multiple trend lines



Figure 8-17. Scatterplot with a single trend line

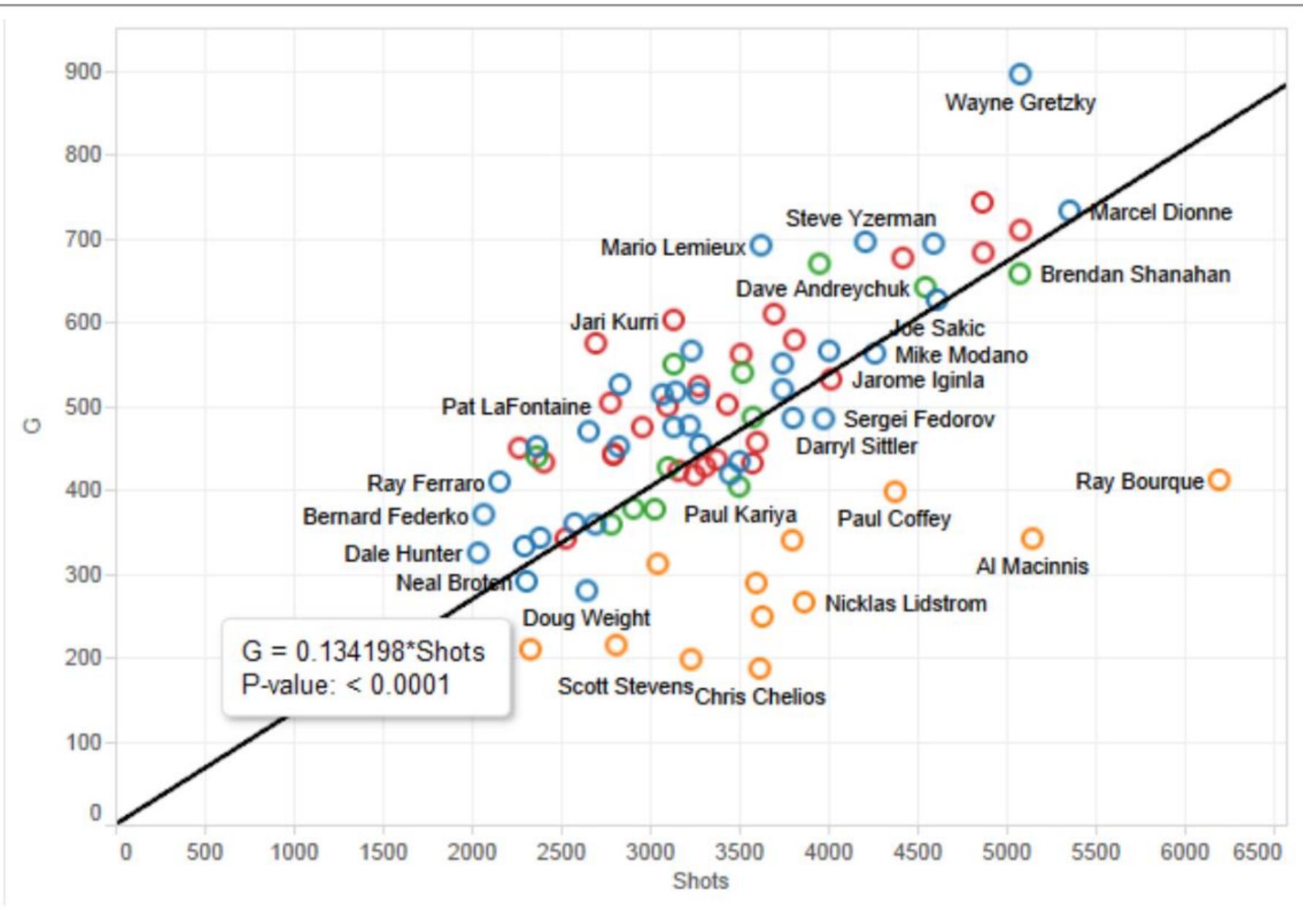


Figure 8-18. Equation and p-value of trend line



# Proportions and Percentages

1

Scatterplots

2

Stacked Bars

3

Regression and Trend Lines

4

The Quadrant Chart



# The Quadrant Chart

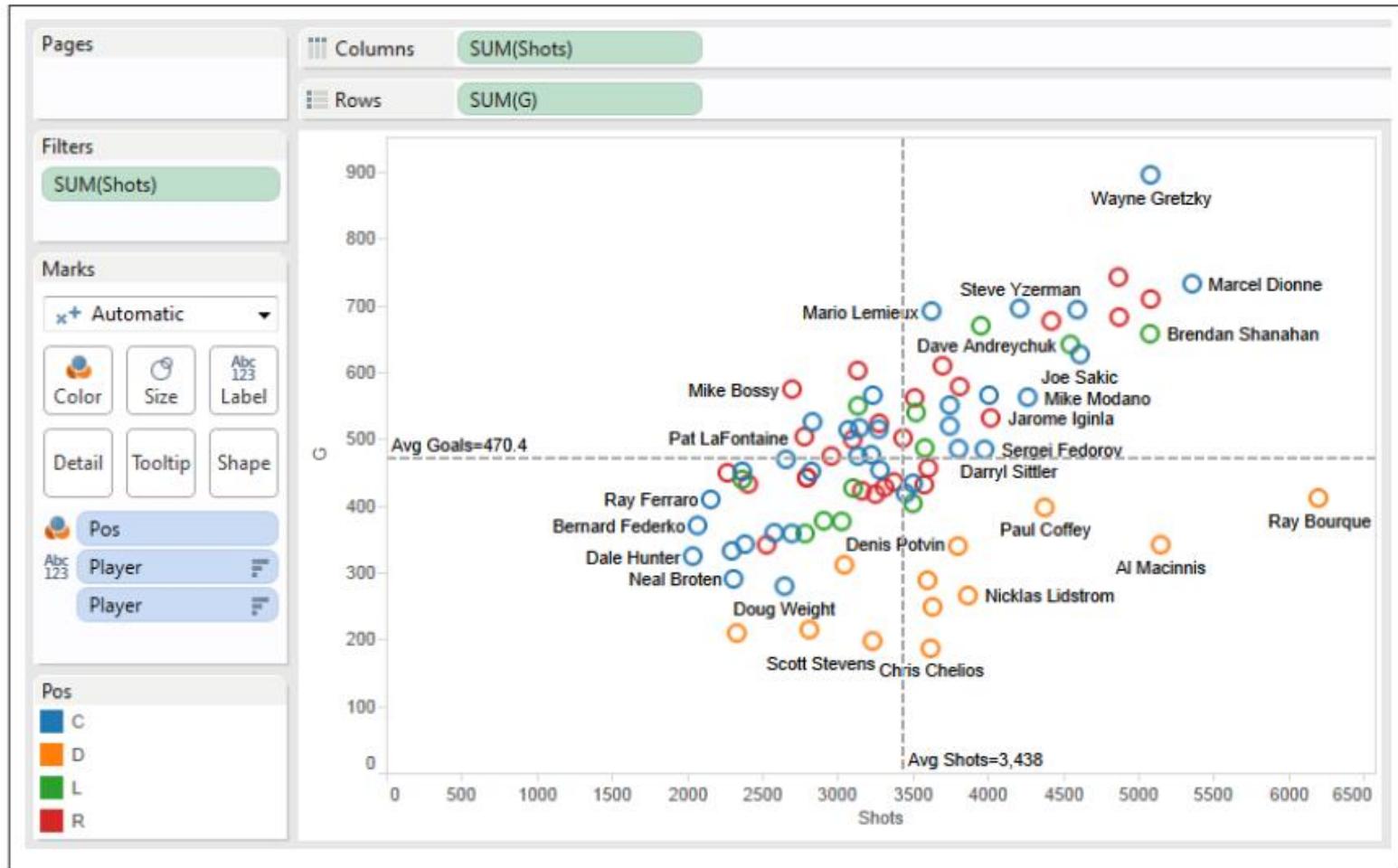


Figure 8-21. The four quadrant scatterplot

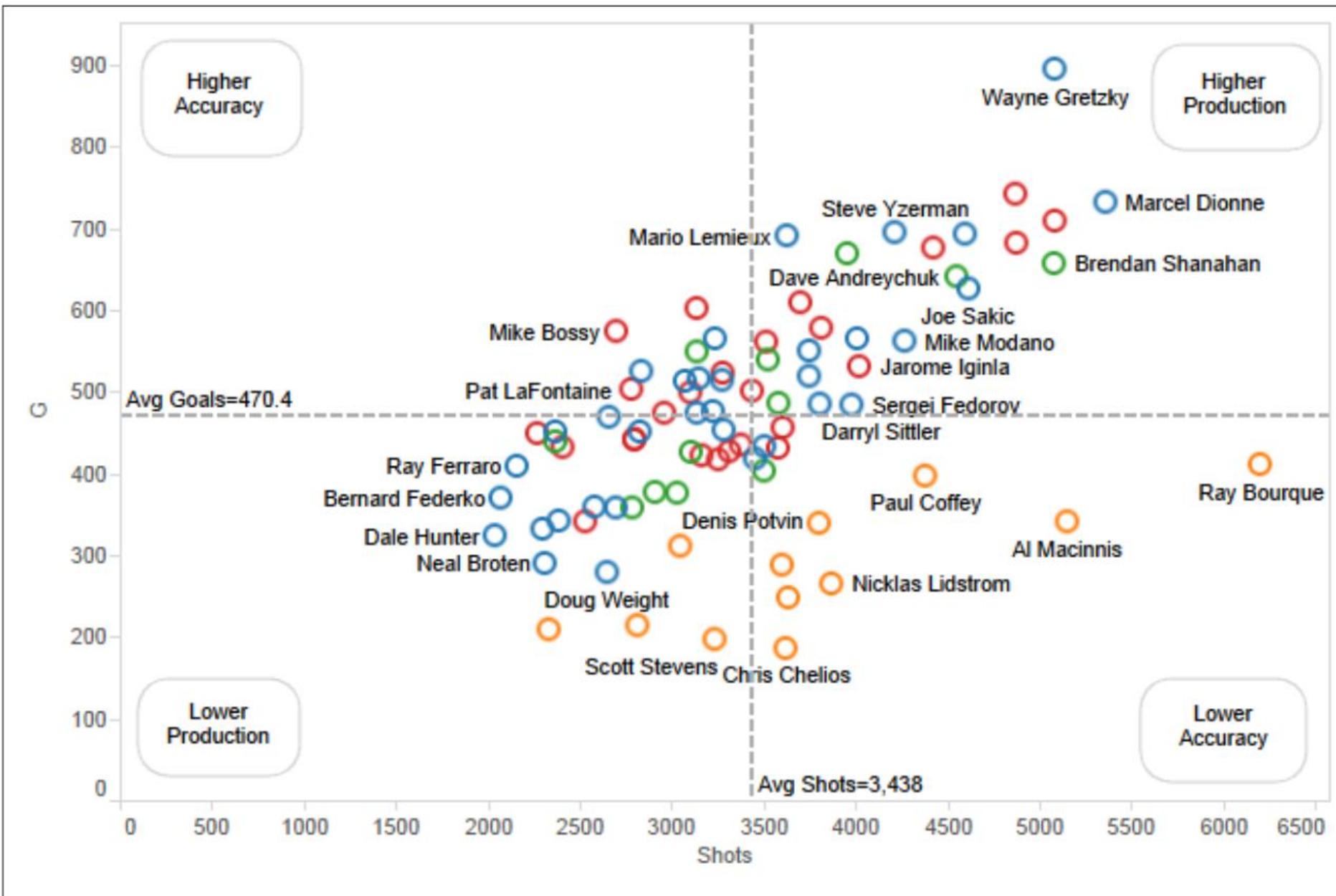


Figure 8-22. The completed quadrant scatterplot



# Proportions and Percentages

1

Scatterplots

2

Stacked Bars

3

Regression and Trend Lines

4

The Quadrant Chart



# DATA VISUALIZATION



# **UNIT - V (PART - II)**



# Changes overtime

1

The origin of time charts

2

The line chart

3

The dual axis line chart

4

The connected scatterplot



# Changes overtime

5 The date filed type and seasonality

6 The timeline

7 The slope graph



# The origin of time charts

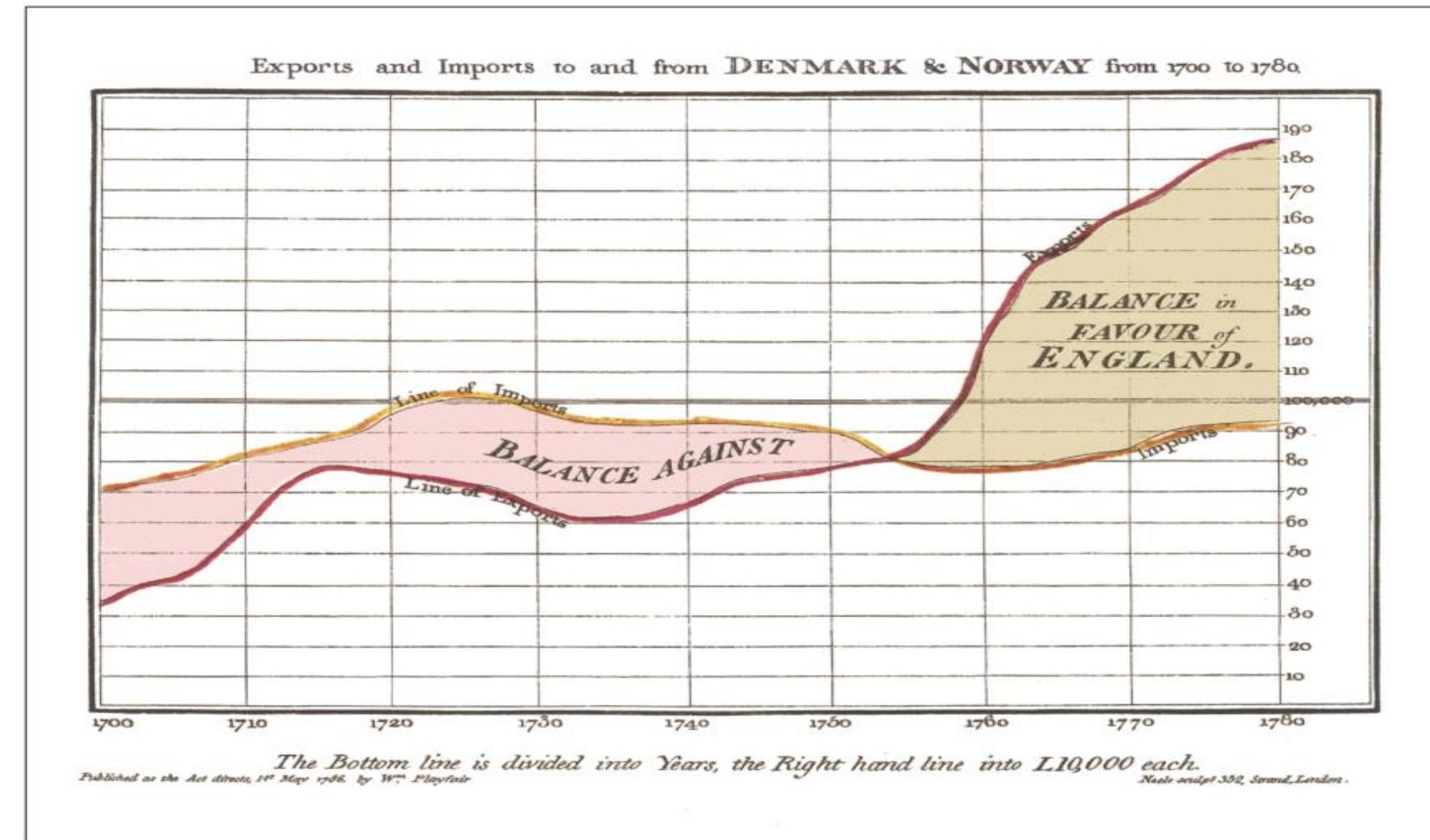


Figure 9-1. The first line chart, by William Playfair in 1786



# The line chart

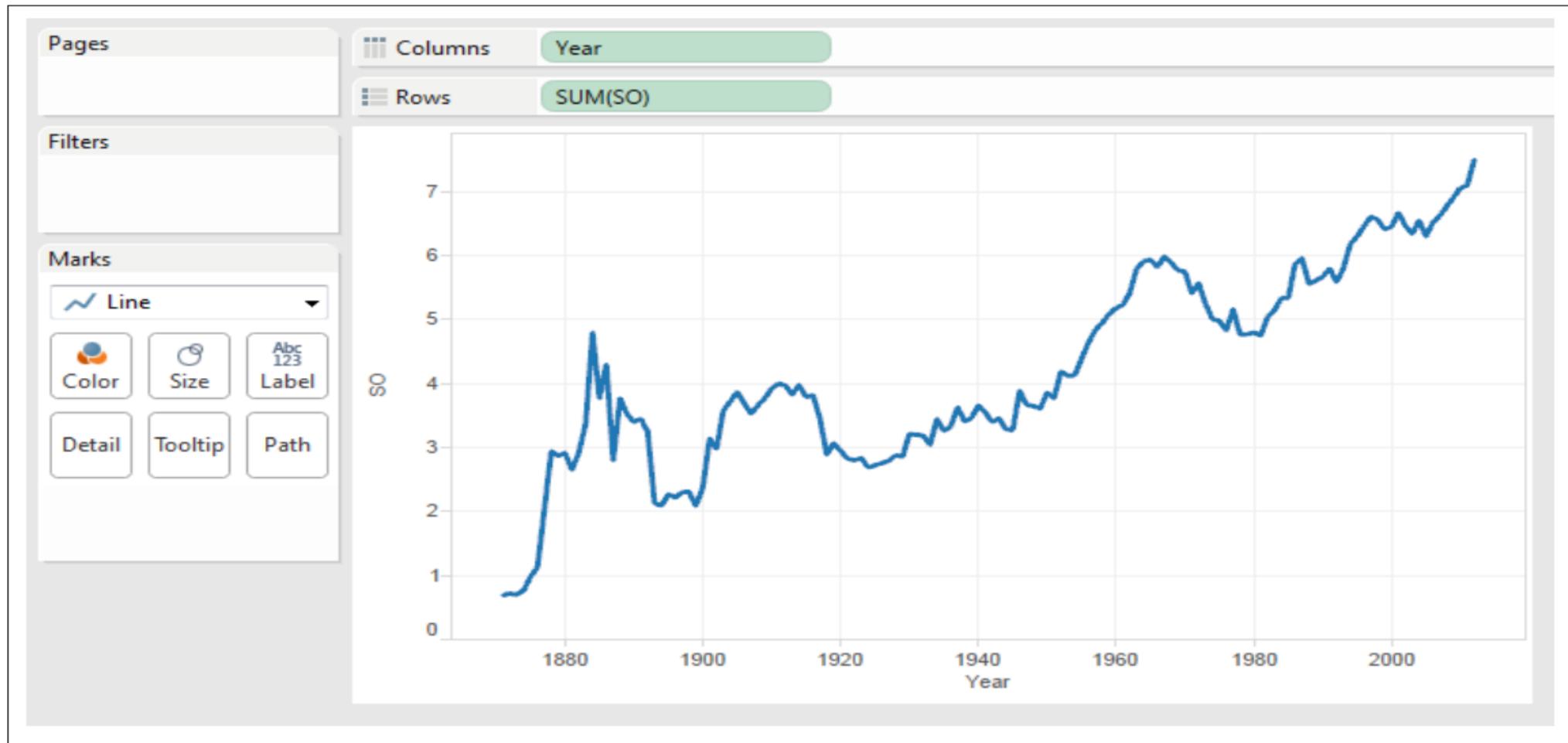


Figure 9-3. A simple line chart showing the increase

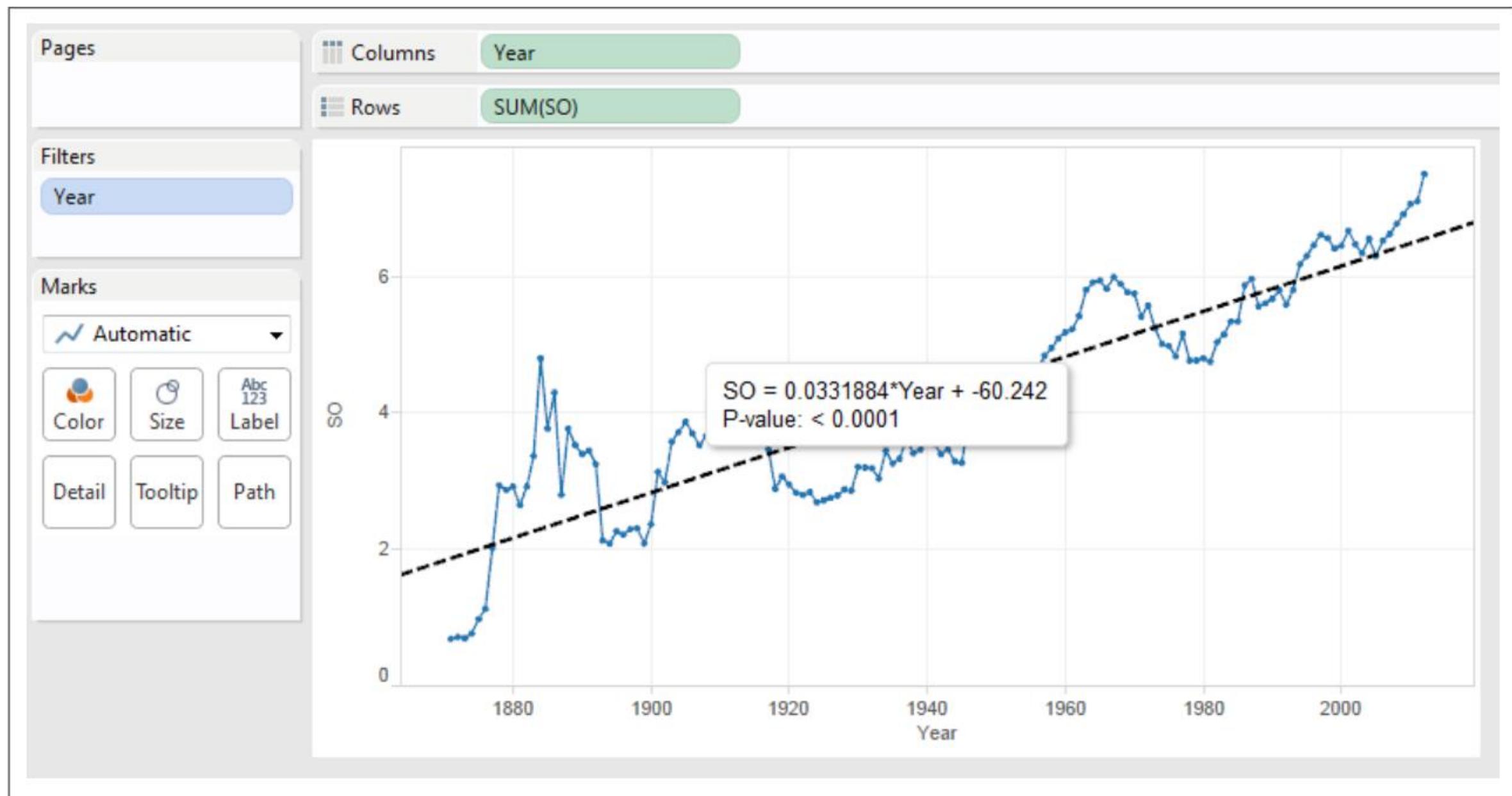


Figure 9-4. Strikeout line chart with trend line added

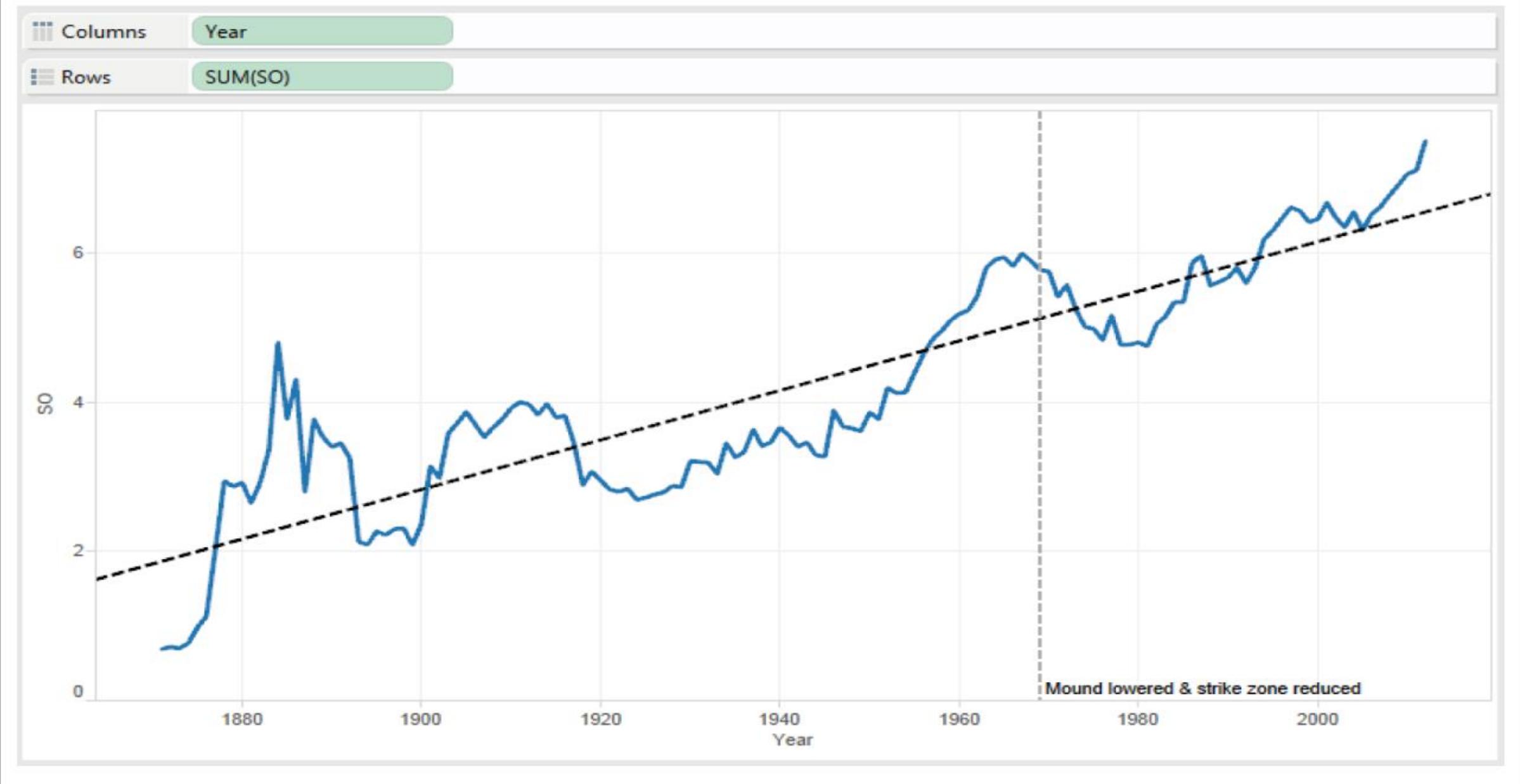


Figure 9-5. Updated line chart with added reference line



# The dual axis line chart

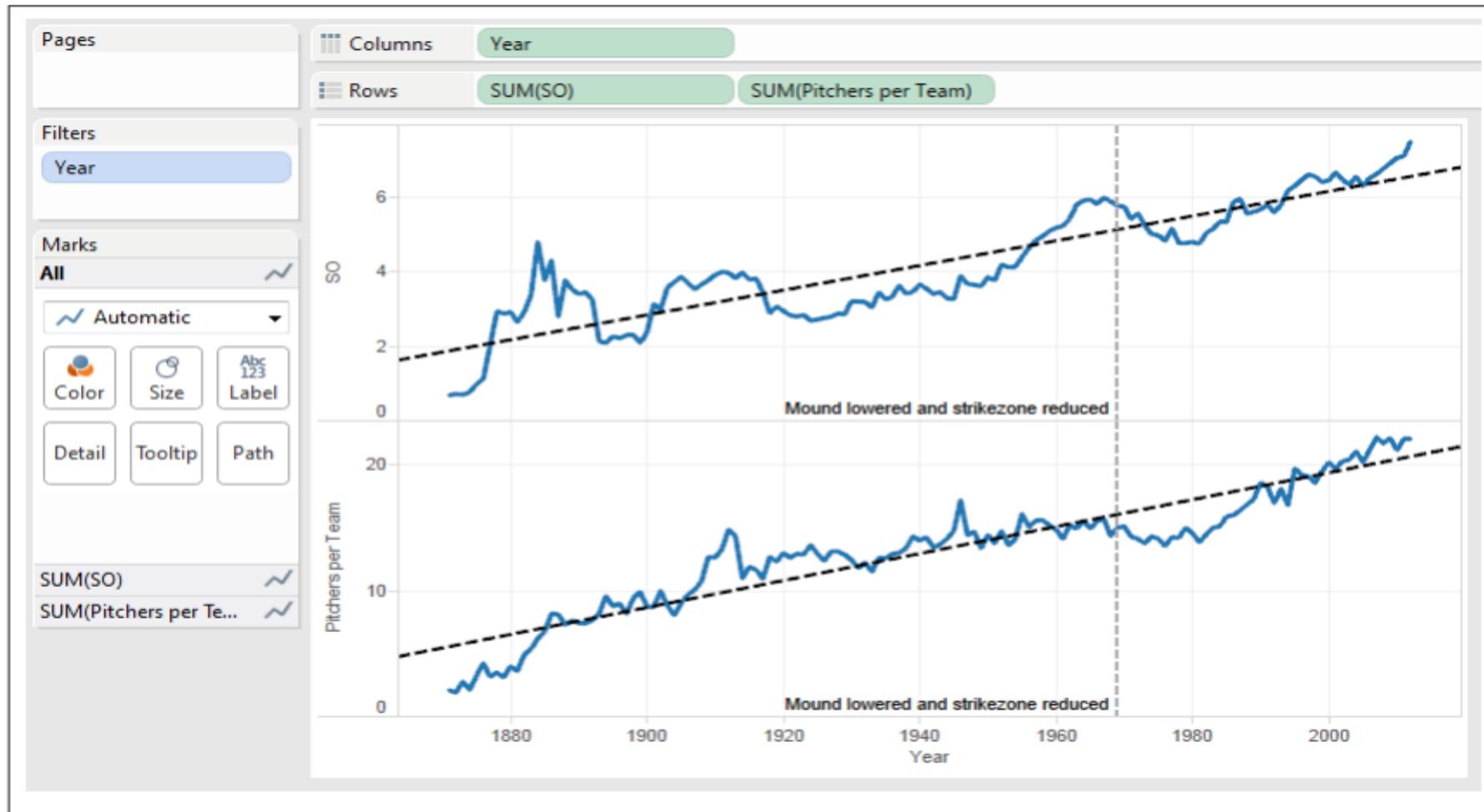


Figure 9-6. Two line charts, shown one above the other

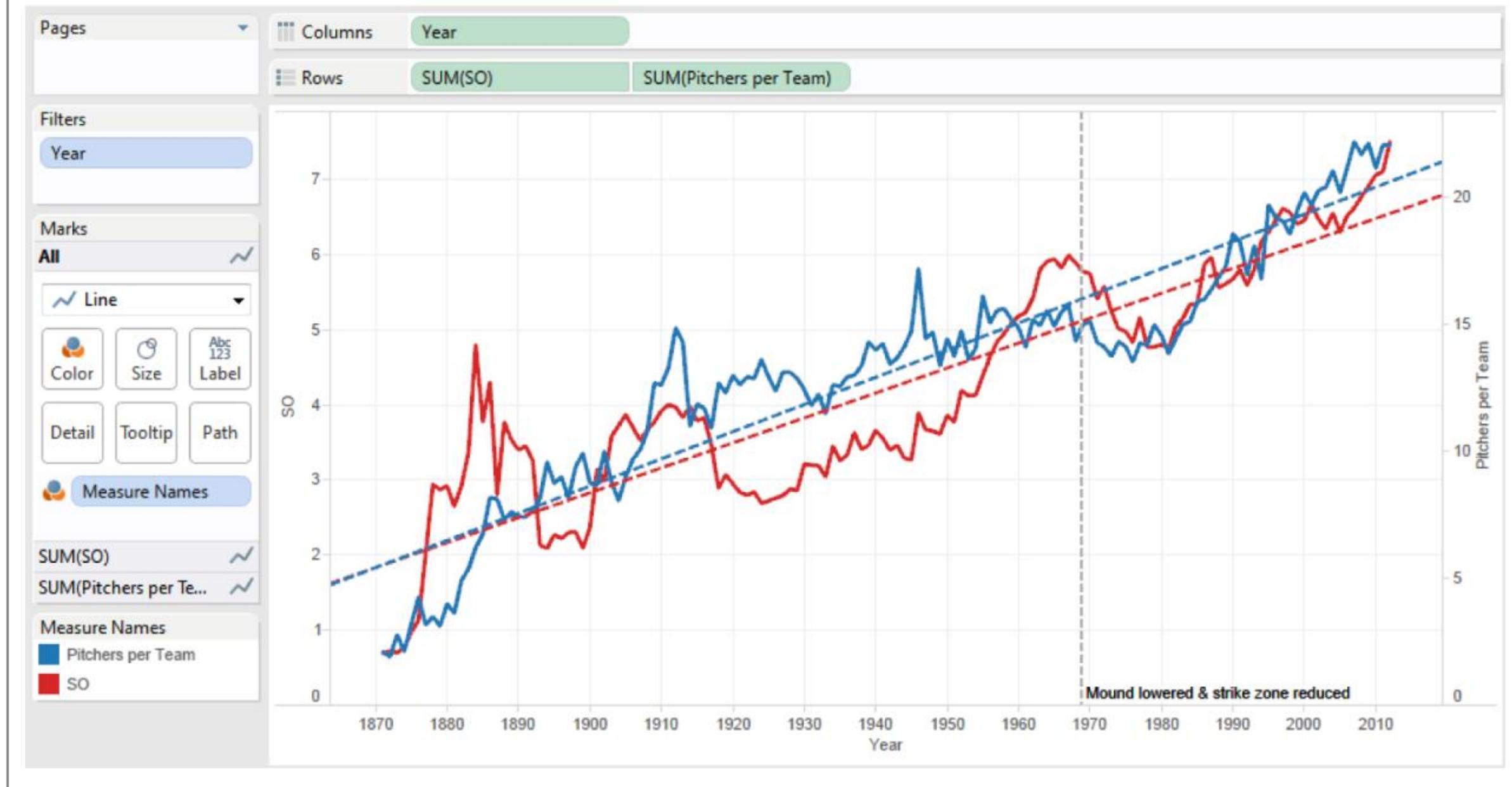


Figure 9-7. The dual-axis line chart



# The connected scatterplot

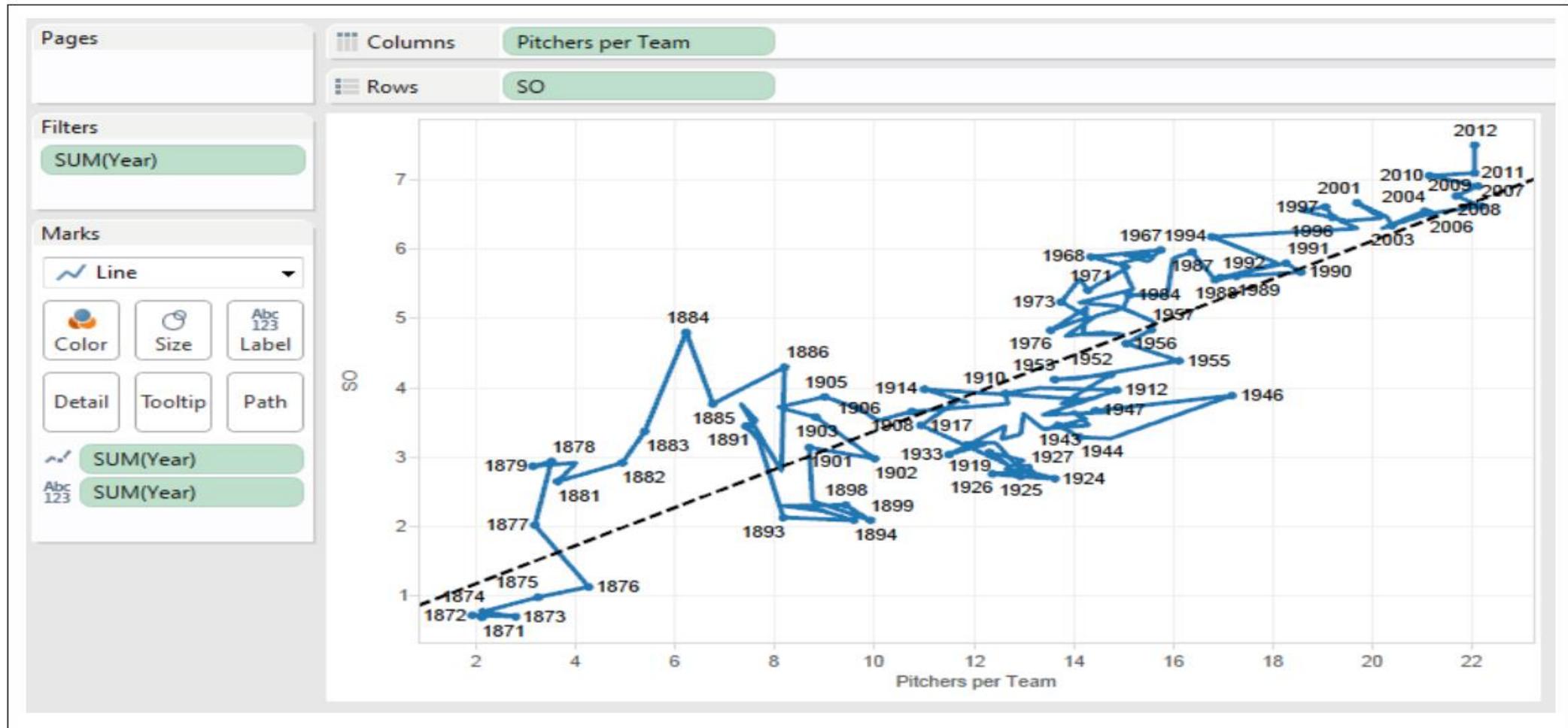


Figure 9-11. The connected scatterplot

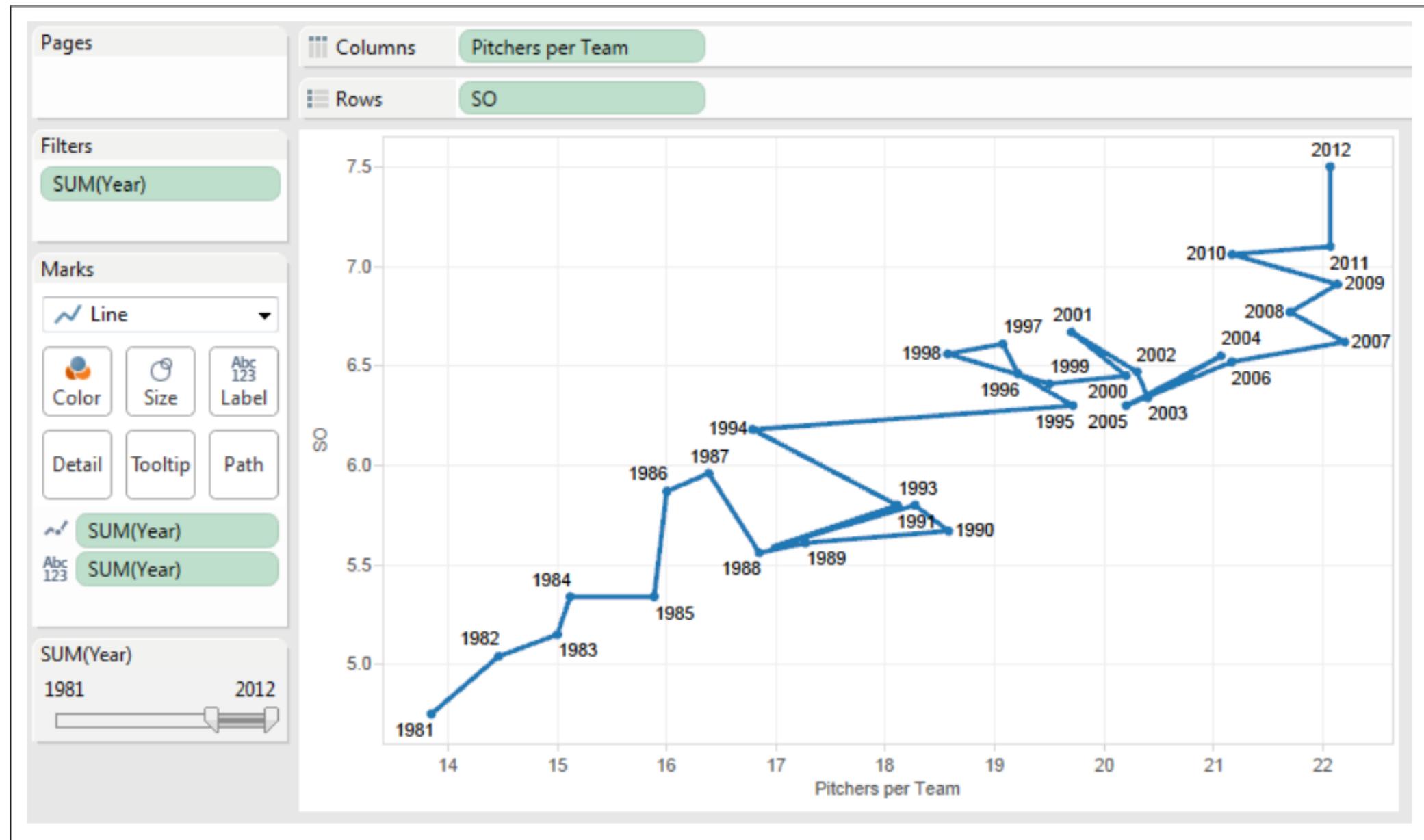


Figure 9-12. A filtered connected scatterplot

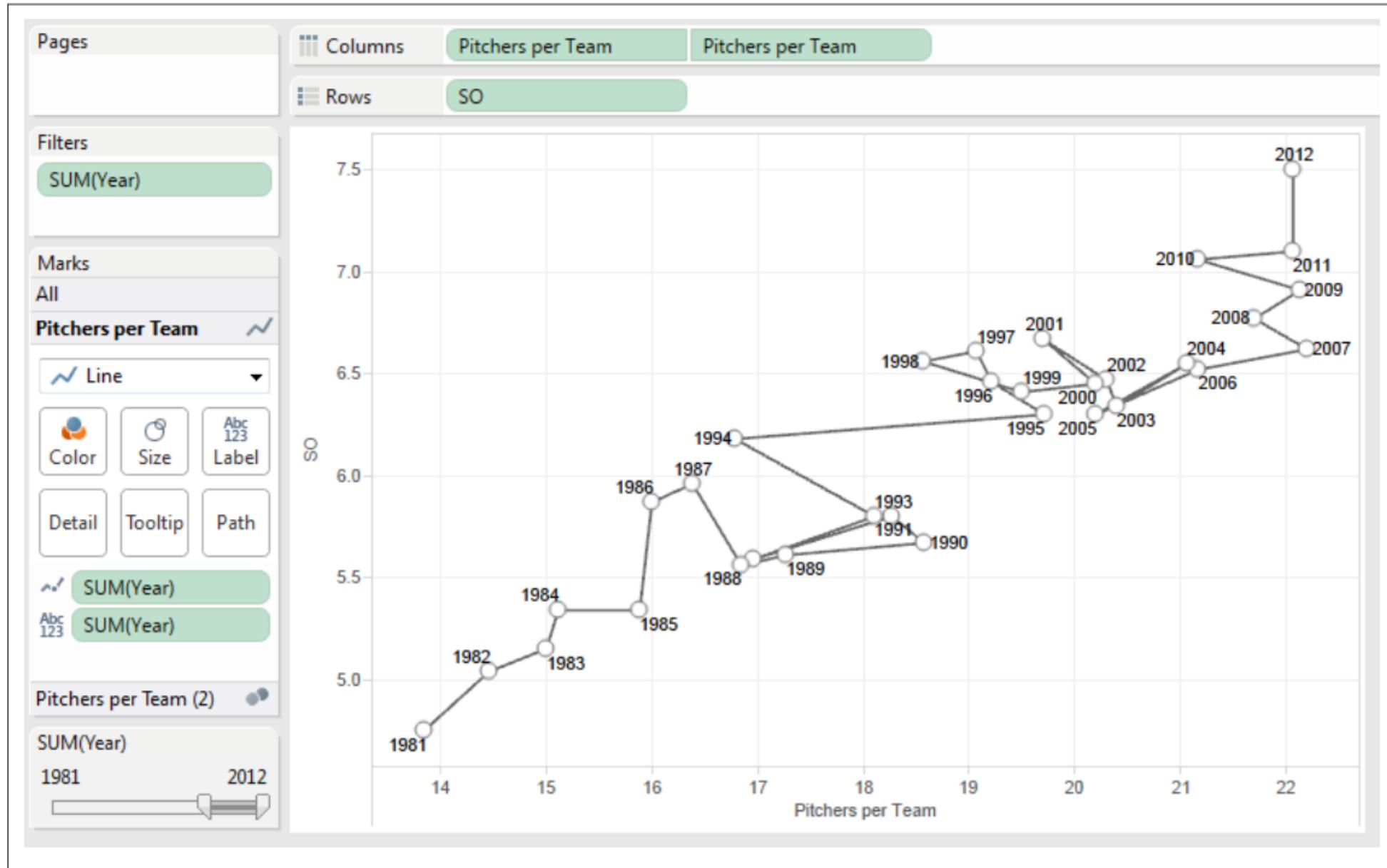


Figure 9-13. The formatted connected scatterplot



# The date filed type and seasonality

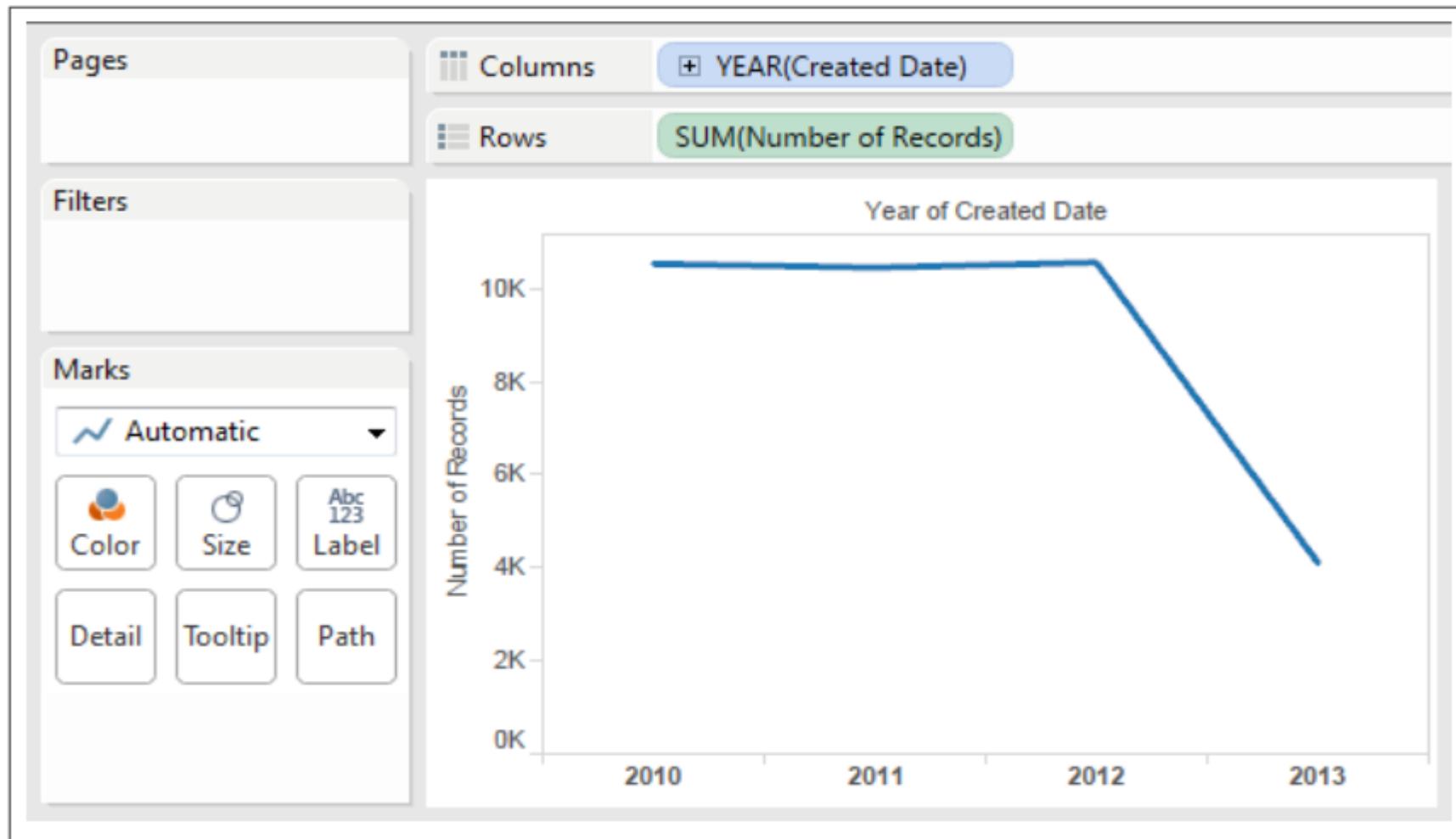


Figure 9-14. Default annual timeline

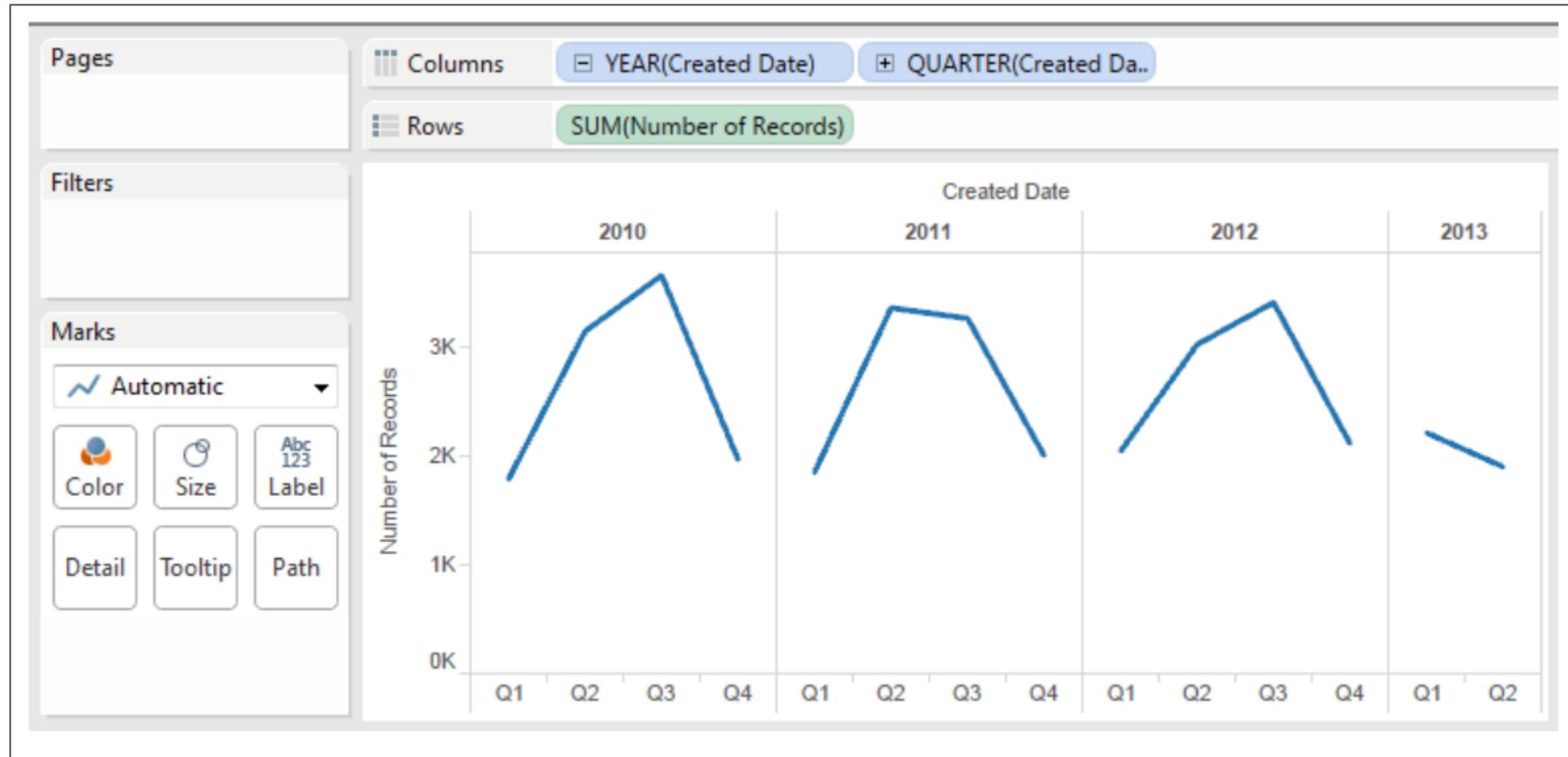
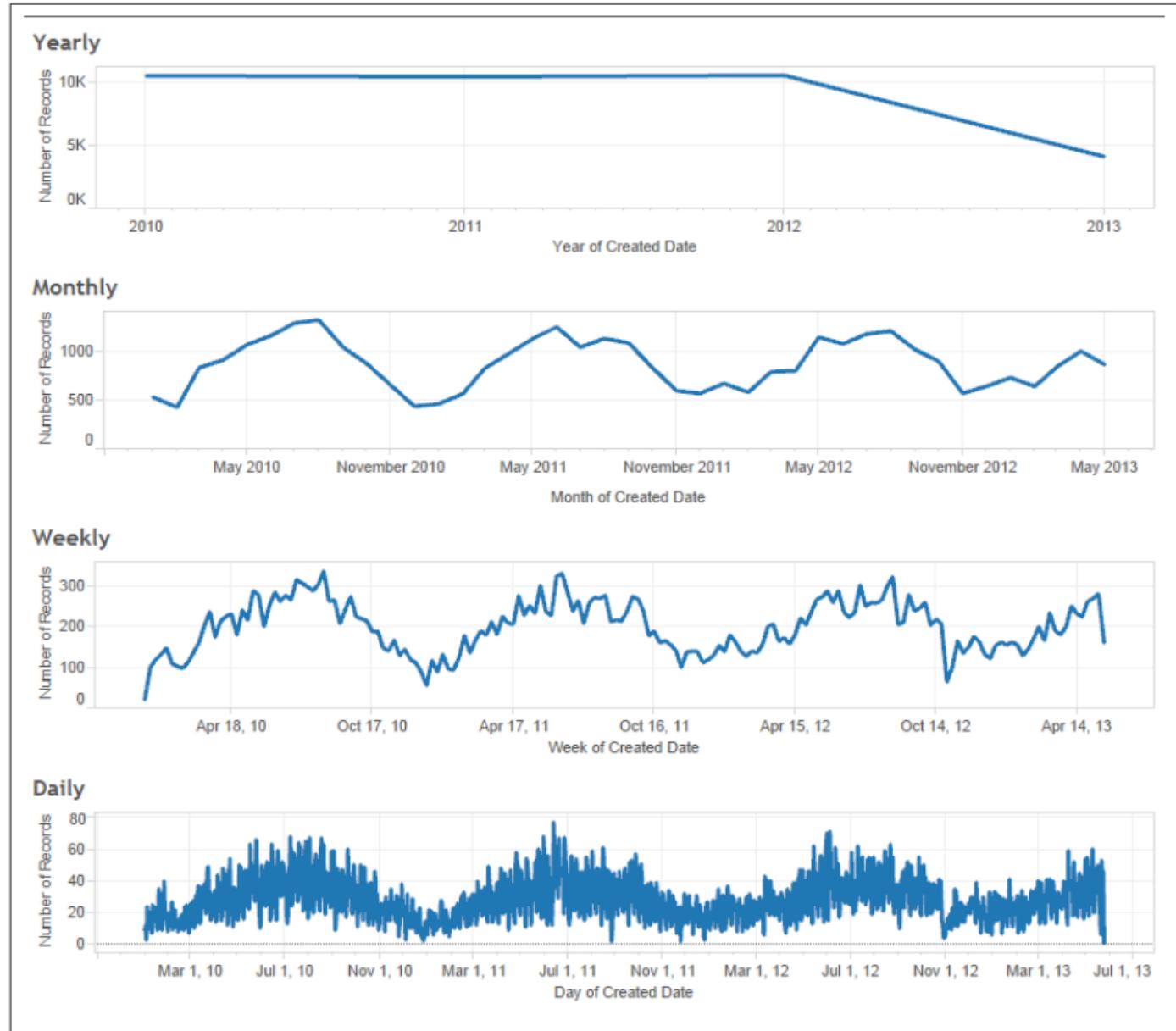


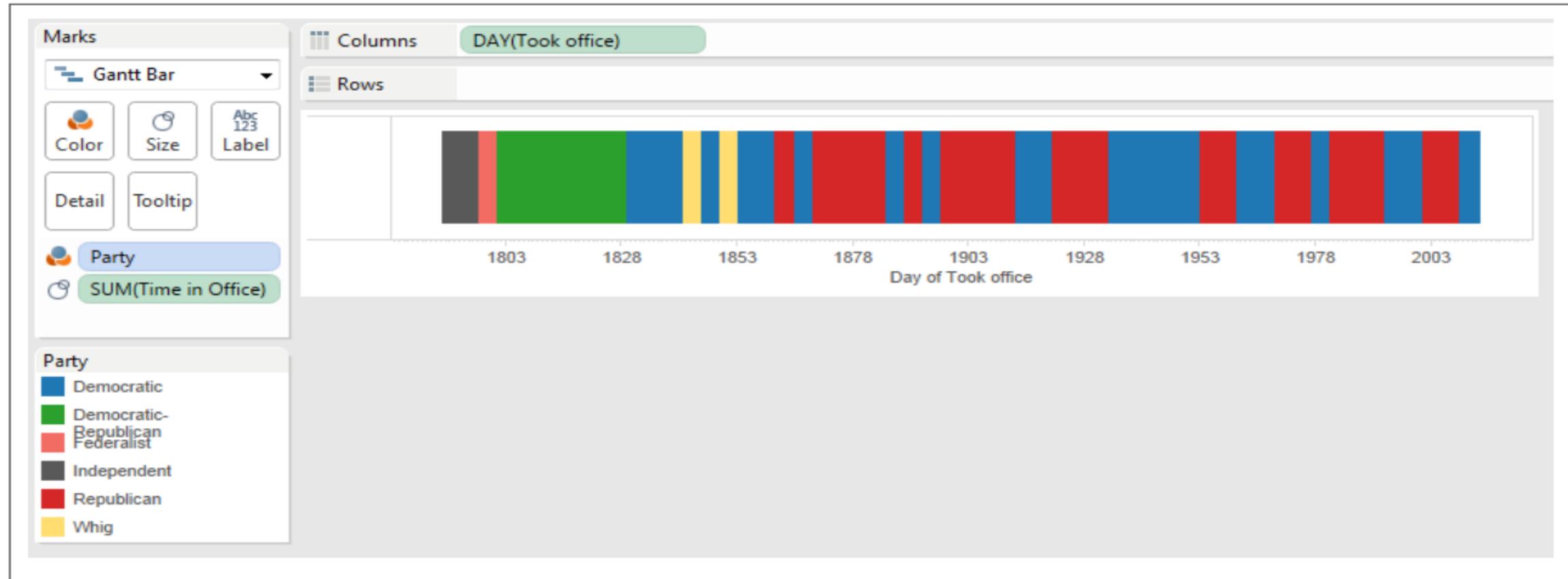
Figure 9-15. Rat sightings by quarter



*Figure 9-16. Rat sightings in New York shown yearly, monthly, weekly, and daily*



# The timeline



*Figure 9-22. A simple timeline of U.S. presidential administrations, colored by political party*



# The slope graph

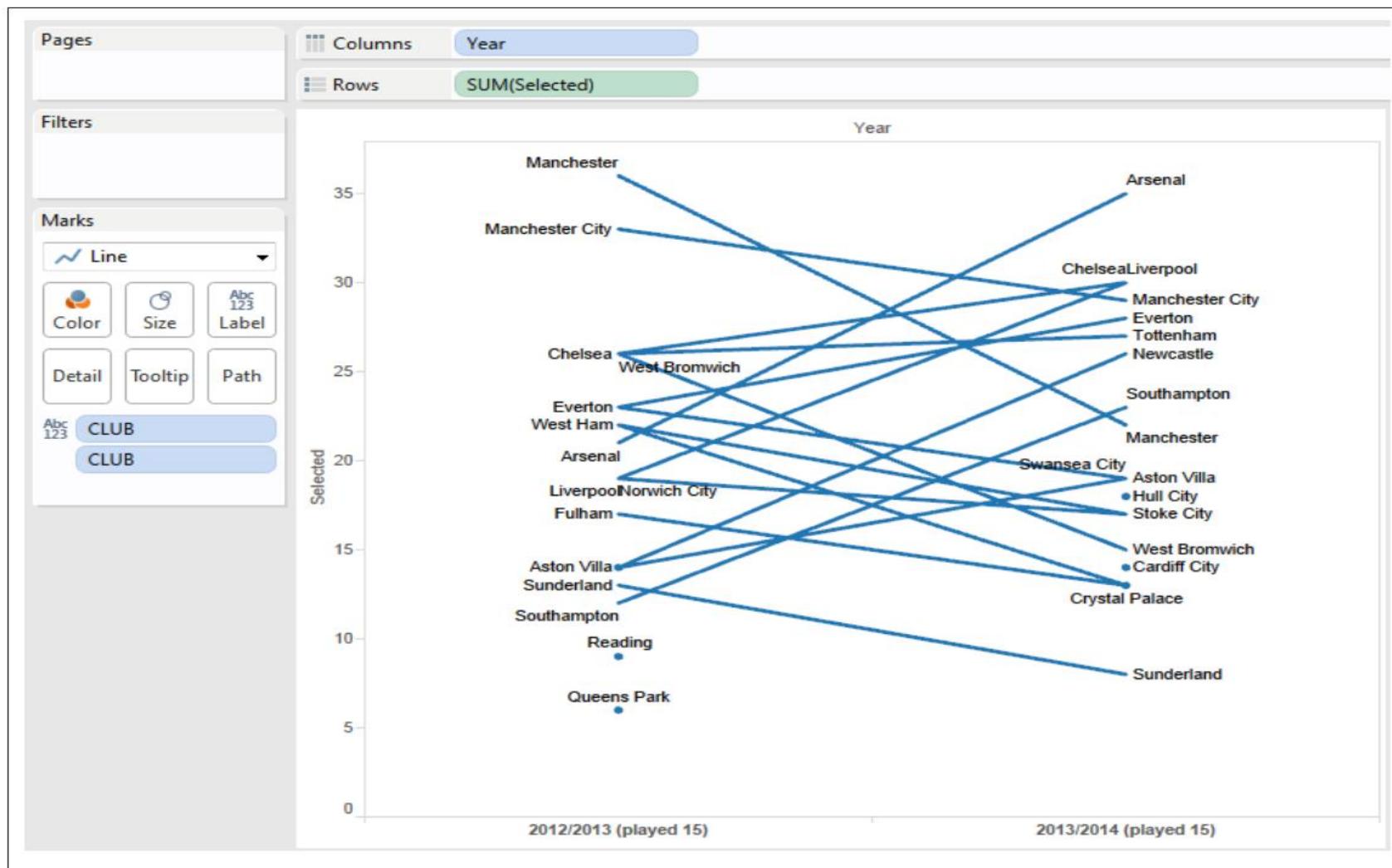


Figure 9-29. Creating a basic slopegraph in Tableau

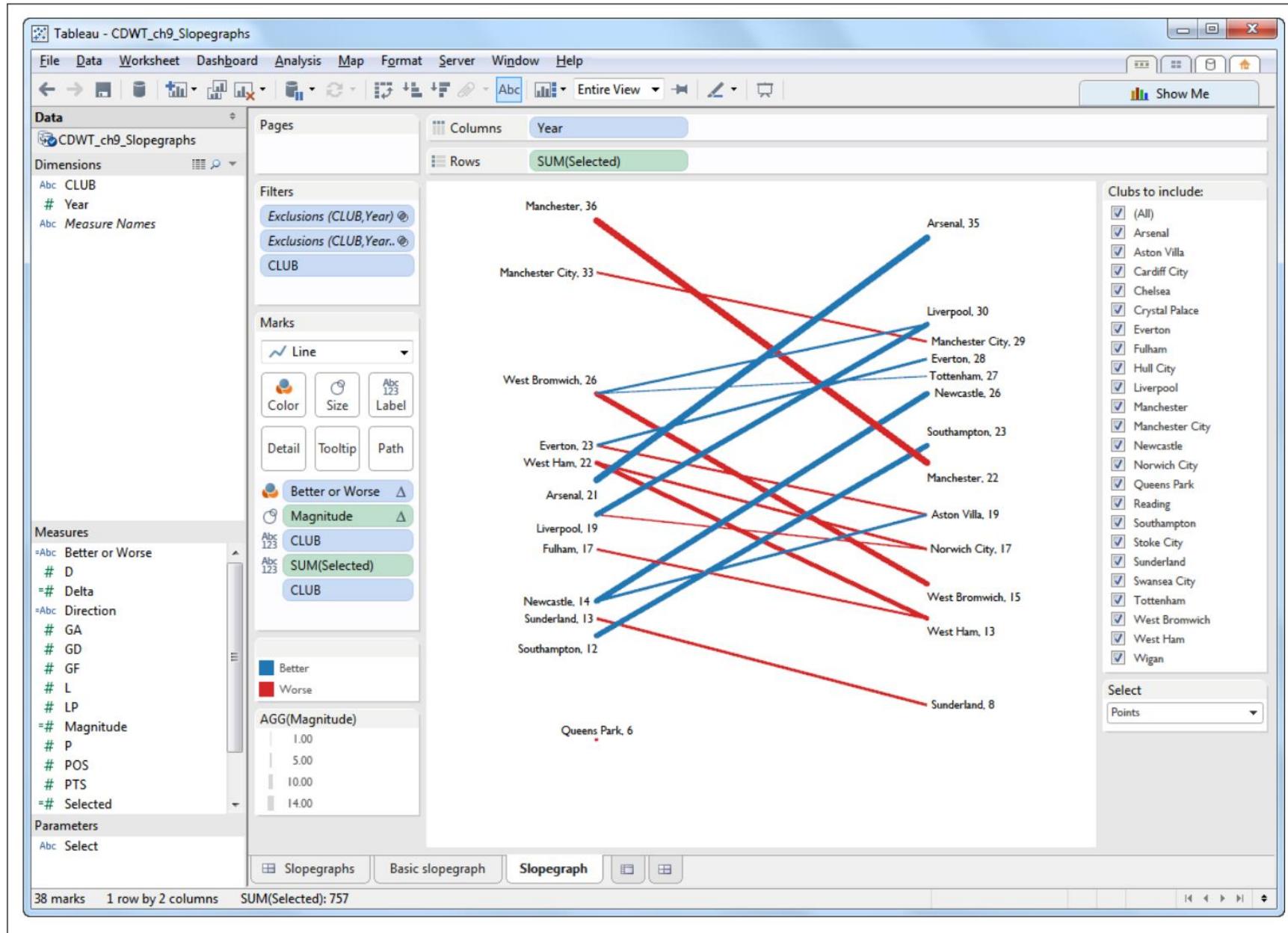


Figure 9-33. An updated slopegraph