wbs
WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

For the Change Makers

Advanced Programming for Data Science

Week 3: Data Visualization
Information Systems and Management
Warwick Business School

Data Visualization

How many times does the digit 7 appear?

5 2 8 3 6 1 9 3 6 2 5 3 7 4 3 8 3

8 5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9

1 0 2 7 5 2 8 3 6 1 6 2 9 3 8 3 8

5 8 4 7 2 0 3 7 3 5 4 7 1 8 2 0 1

2 5 3 6 4 3 9 1 0 8 9 5 7 3 4 5 3

2 7 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8

4 7 2 0 3 7 3 5 4 7 1 8 2 0 1 9 6

2 1 4 4 3 9 3 6 5 2 4 9 1 0 2 7 5

2 8 3 6 1 6 2 9 3 8 3 8 5 8 4 7 2

0 3 7 3 5 4 7 1 8 2 0 1 2 5 3 6 4

3 9 1 0 8 9 5 7 3 4 5 3 2 7 5 2 8

3 6 1 6 2 4 6 2 7 5 9 1 5 2 6 3 6

5 2 8 3 6 1 9 3 6 2 5 3 7 4 3 8 3

8 5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9

1 0 2 7 5 2 8 3 6 1 6 2 9 3 8 3 8

5 8 4 7 2 0 3 7 3 5 4 7 1 8 2 0 1

2 5 3 6 4 3 9 1 0 8 9 5 7 3 4 5 3

2 7 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8

4 7 2 0 3 7 3 5 4 7 1 8 2 0 1 9 6

2 1 4 4 3 9 3 6 5 2 4 9 1 0 2 7 5

2 8 3 6 1 6 2 9 3 8 3 8 5 8 4 7 2

0 3 7 3 5 4 7 1 8 2 0 1 2 5 3 6 4

3 9 1 0 8 9 5 7 3 4 5 3 2 7 5 2 8

3 6 1 6 2 4 6 2 7 5 9 1 5 2 6 3 6

5 2 8 3 6 1 9 3 6 2 5 3 7 4 3 8 3

8 5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9

1 0 2 7 5 2 8 3 6 1 6 2 9 3 8 3 8

5 8 4 7 2 0 3 7 3 5 4 7 1 8 2 0 1

2 5 3 6 4 3 9 1 0 8 9 5 7 3 4 5 3

2 7 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8

4 7 2 0 3 7 3 5 4 7 1 8 2 0 1 9 6

2 1 4 4 3 9 3 6 5 2 4 9 1 0 2 7 5

2 8 3 6 1 6 2 9 3 8 3 8 5 8 4 7 2

0 3 7 3 5 4 7 1 8 2 0 1 2 5 3 6 4

3 9 1 0 8 9 5 7 3 4 5 3 2 7 5 2 8

3 6 1 6 2 4 6 2 7 5 9 1 5 2 6 3 6

5 2 8 3 6 1 9 3 6 2 5 3 7 4 3 8 3

8 5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9

1 0 2 7 5 2 8 3 6 1 6 2 9 3 8 3 8

5 8 4 7 2 0 3 7 3 5 4 7 1 8 2 0 1

2 5 3 6 4 3 9 1 0 8 9 5 7 3 4 5 3

2 7 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8

4 7 2 0 3 7 3 5 4 7 1 8 2 0 1 9 6

2 1 4 4 3 9 3 6 5 2 4 9 1 0 2 7 5

2 8 3 6 1 6 2 9 3 8 3 8 5 8 4 7 2

0 3 7 3 5 4 7 1 8 2 0 1 2 5 3 6 4

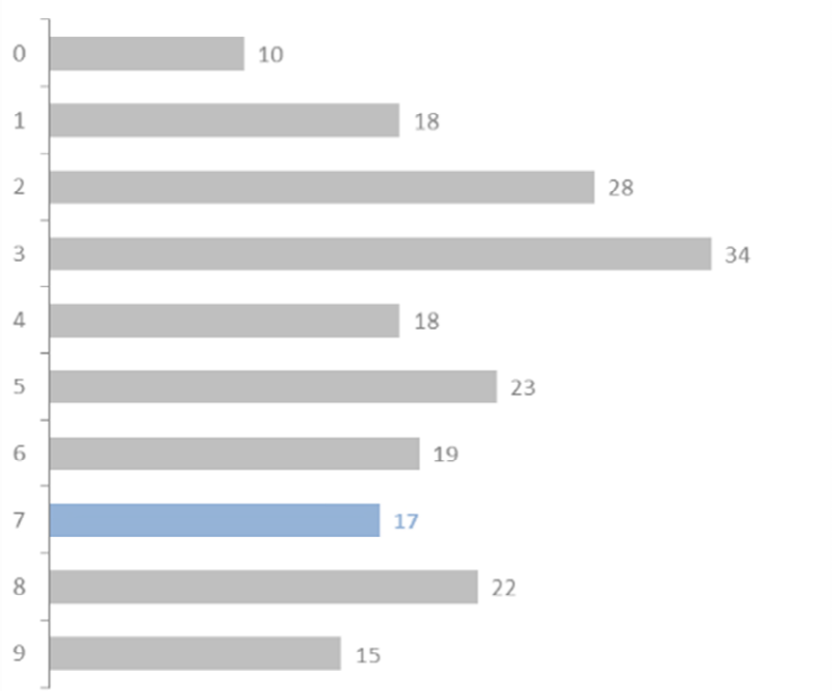
3 9 1 0 8 9 5 7 3 4 5 3 2 7 5 2 8

3 6 1 6 2 4 6 2 7 5 9 1 5 2 6 3 6

7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6

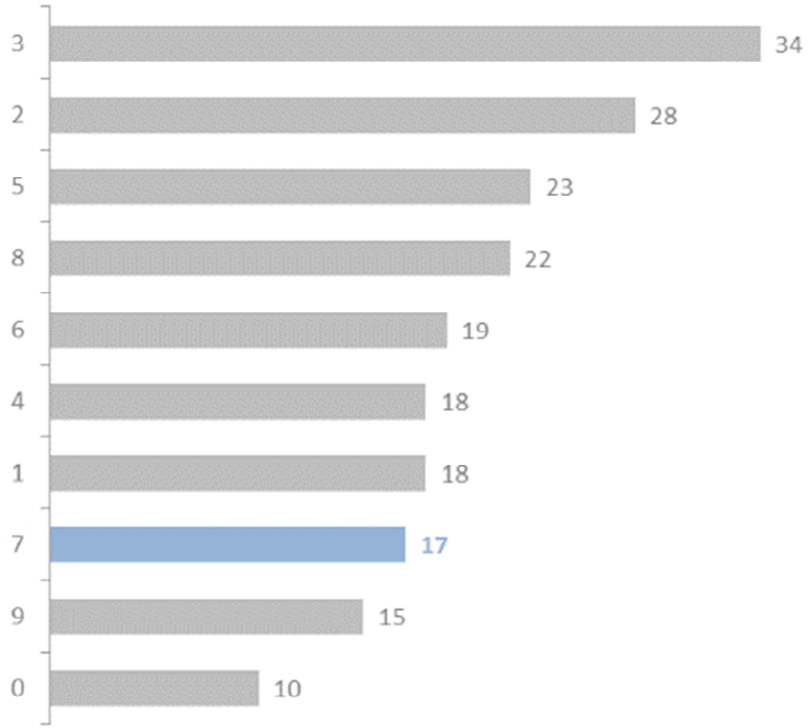
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8	
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1	
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8	
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4	
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6	
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4	
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2	
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3	
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5	
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2	
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6	

of times digit 7 appears: 17



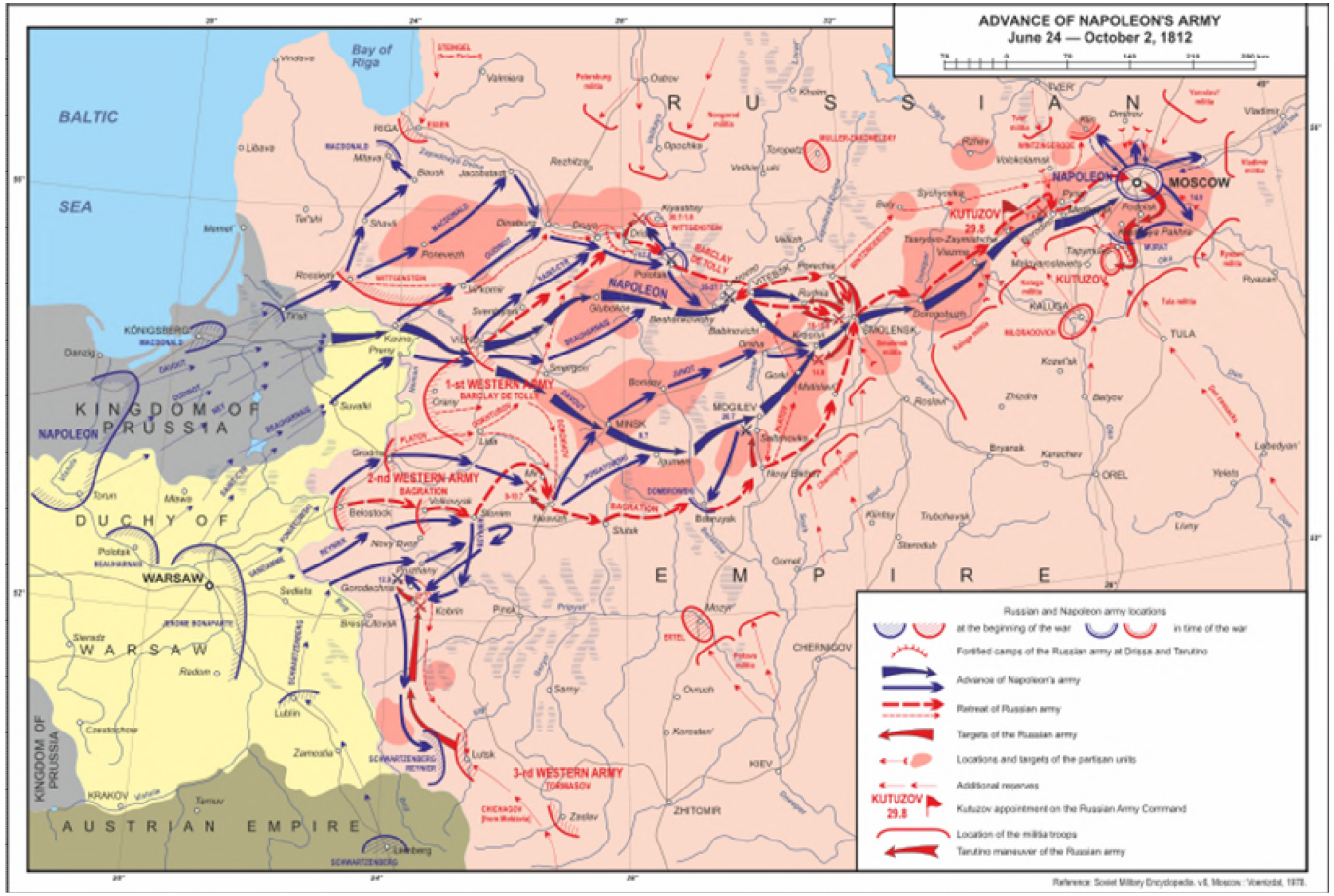
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6

of times digit 7 appears: 17



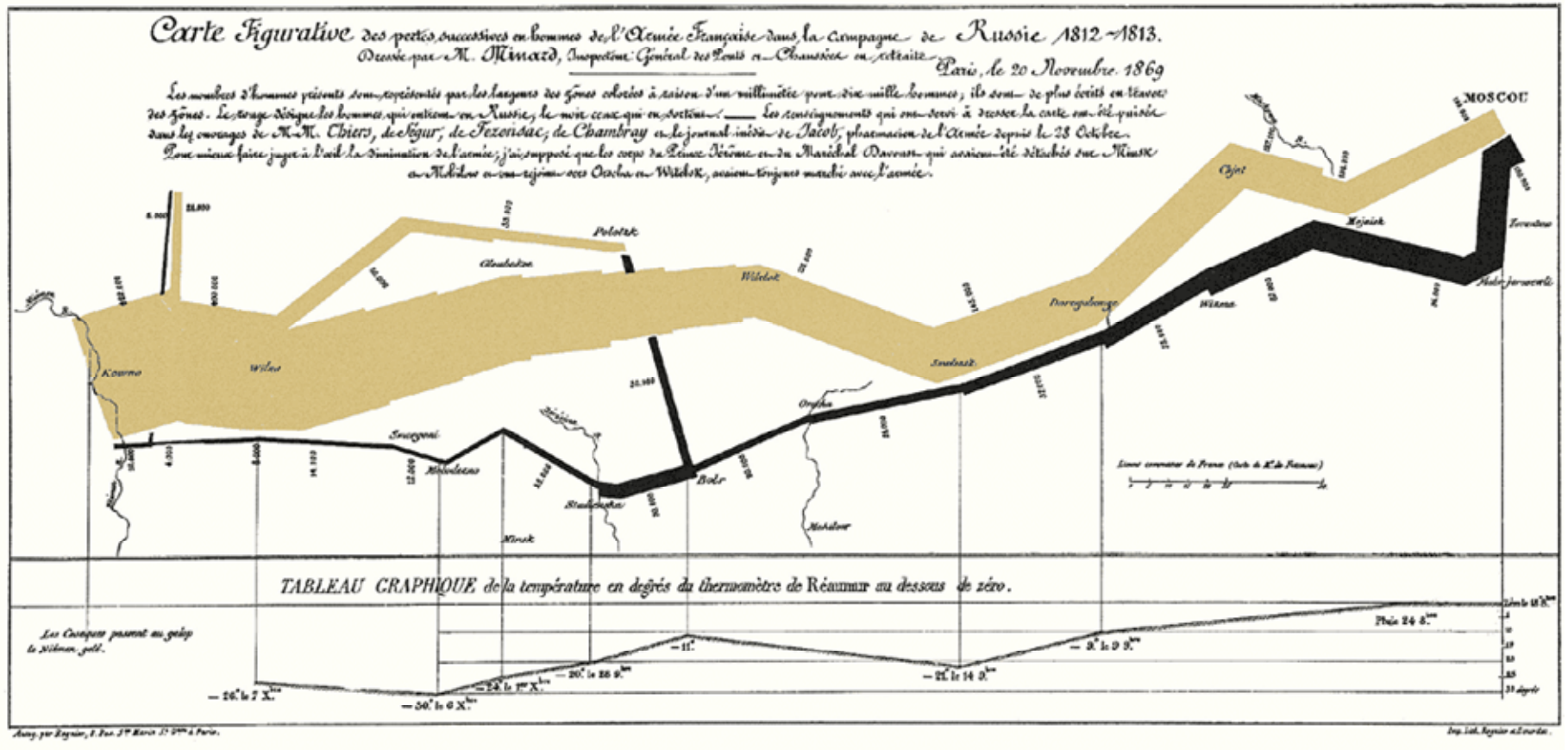
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6

What makes a good chart?



Wikipedia: Patriotic War of 1812
http://en.wikipedia.org/wiki/File:Patriotic_War_of_1812_ENG_map1.svg

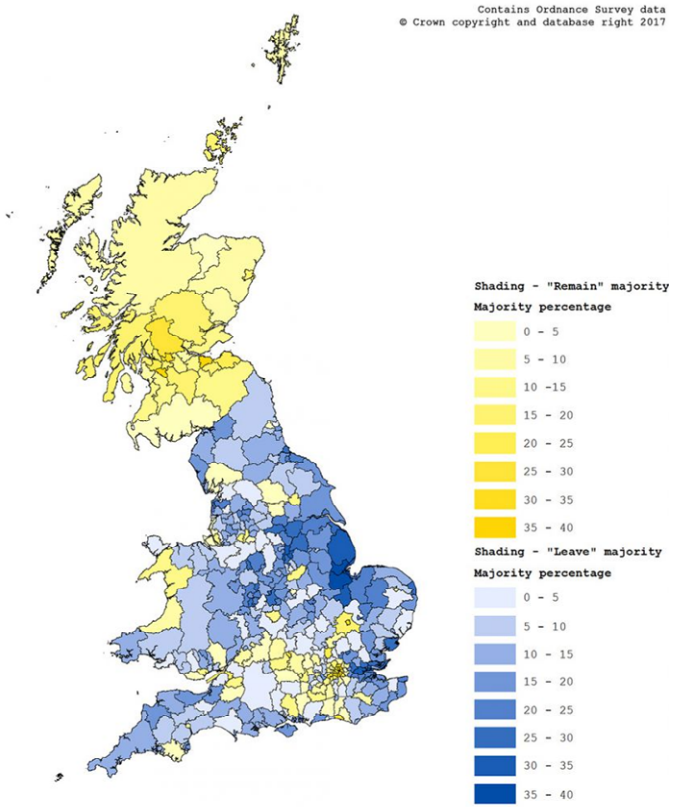
What makes a good chart?



Minard's map of Napoleon's campaign into Russia, 1869
Reprinted in Tufte (2009), p. 41

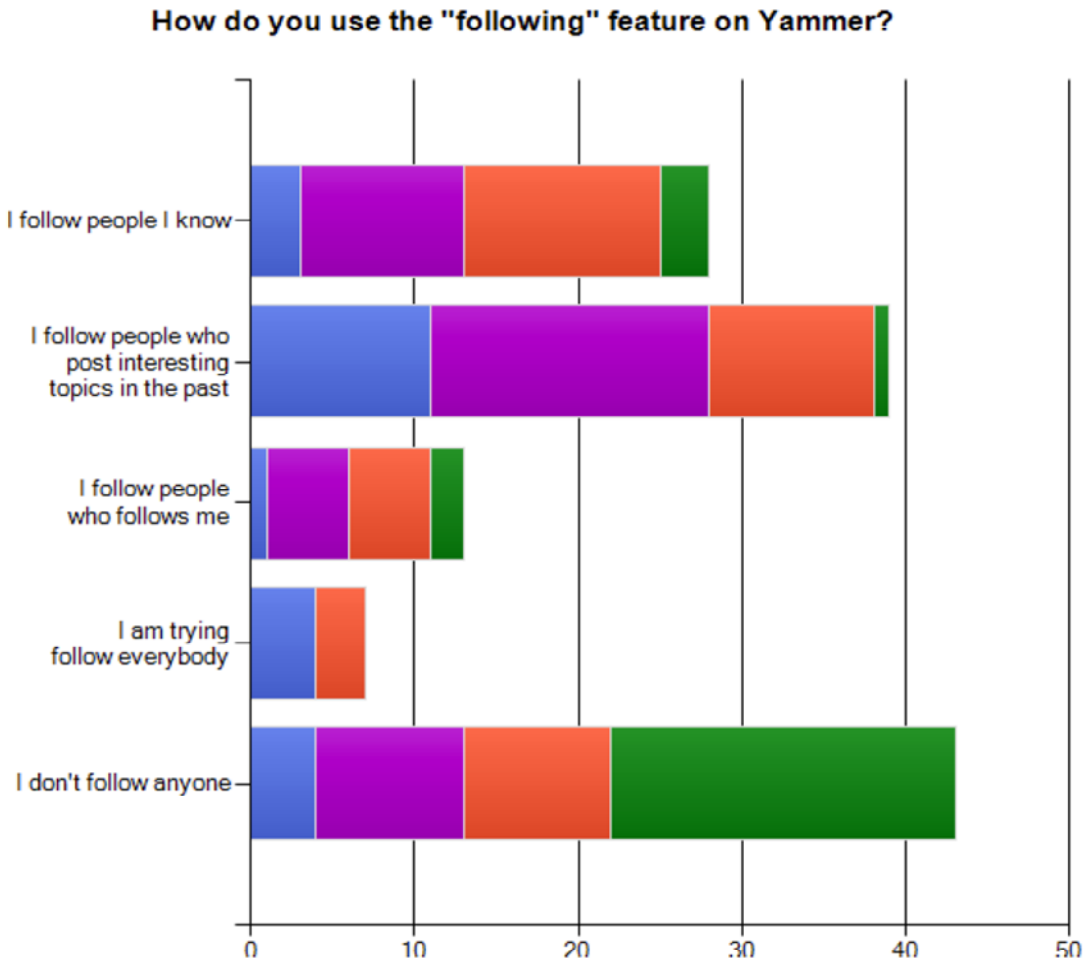
Video Games do this well…

What can you learn from this map?



http://www.ox.ac.uk/news-and-events/oxford-and-brexit/brexit-analysis/mapping-brexit-vote

What makes a good chart?



This is from an academic conference paper.

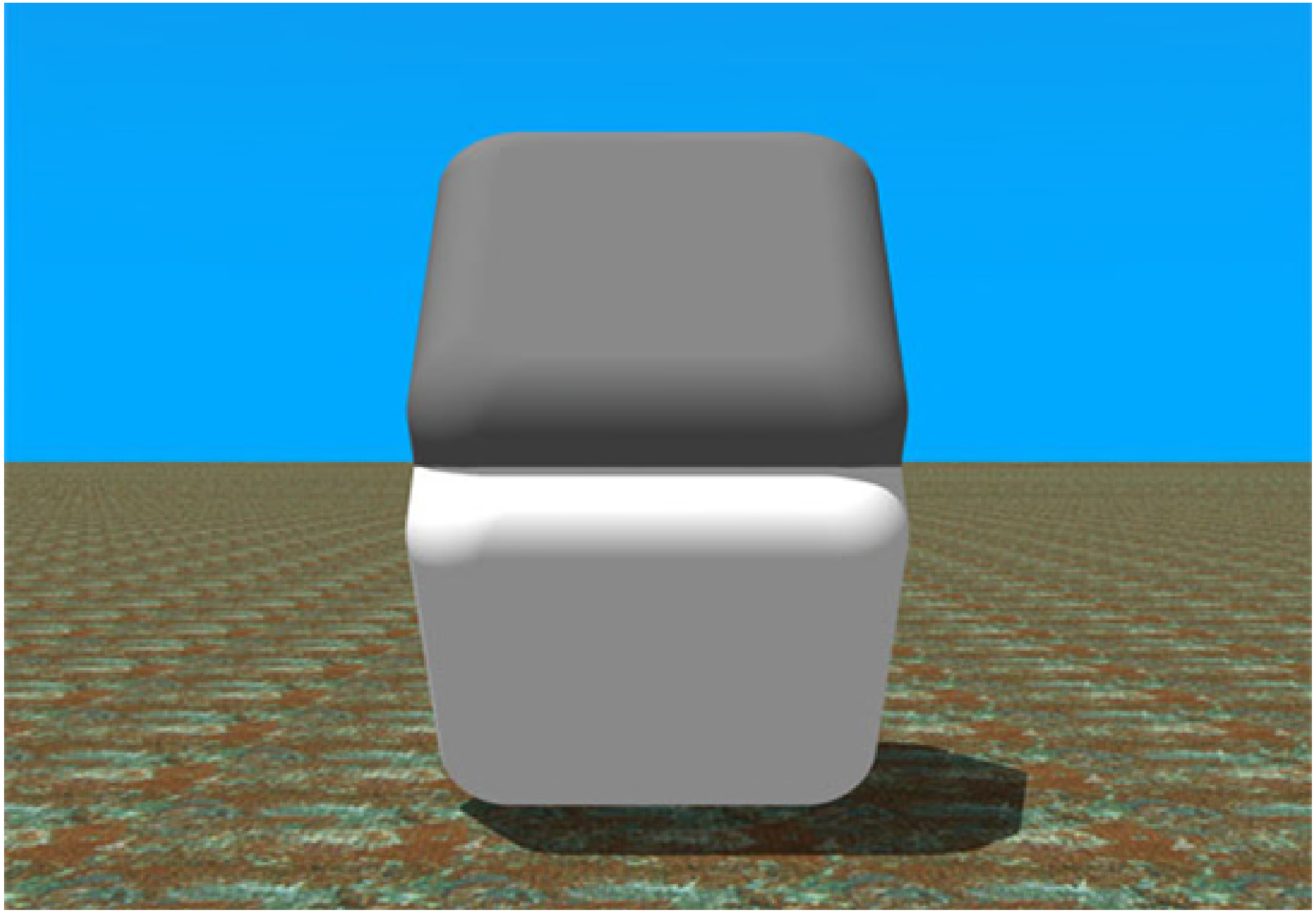
What are the problems with this chart?

Zhang et al. (2010), "A case study of micro-blogging in the enterprise: use, value, and related issues," Proceedings of the 28th International Conference on Human Factors in Computing Systems.

Some basic principles (adapted from Tufte 2009)

~~**How to make a good chart**~~

How not to make a bad chart



Used by Permission of Dr. Beau Lotto (www.LottoLab.org)

Some basic principles (adapted from Tufte 2009)

1. The chart should tell a story
2. The chart should have graphical integrity
3. The chart should minimize graphical complexity

Tufte's fundamental principle:
Above all else show the data

Principle 1: The chart should tell a story

Graphics should be clear on their own

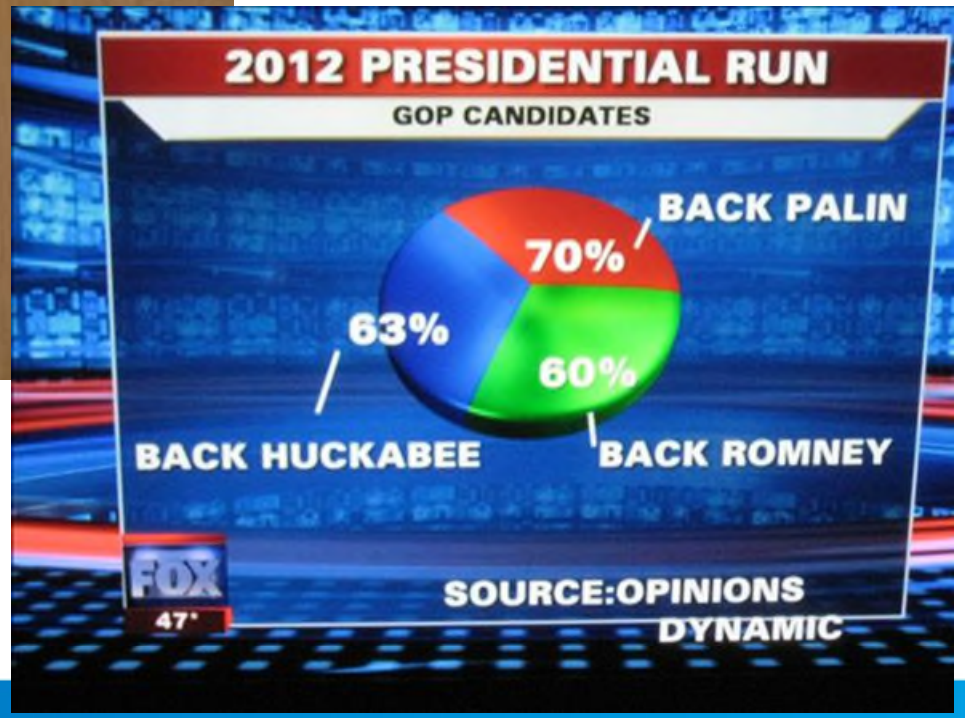
The depictions should enable meaningful comparison

The chart should yield insight beyond the text

“If the statistics are boring, then you’ve got the wrong numbers.” (Tufte 2009)

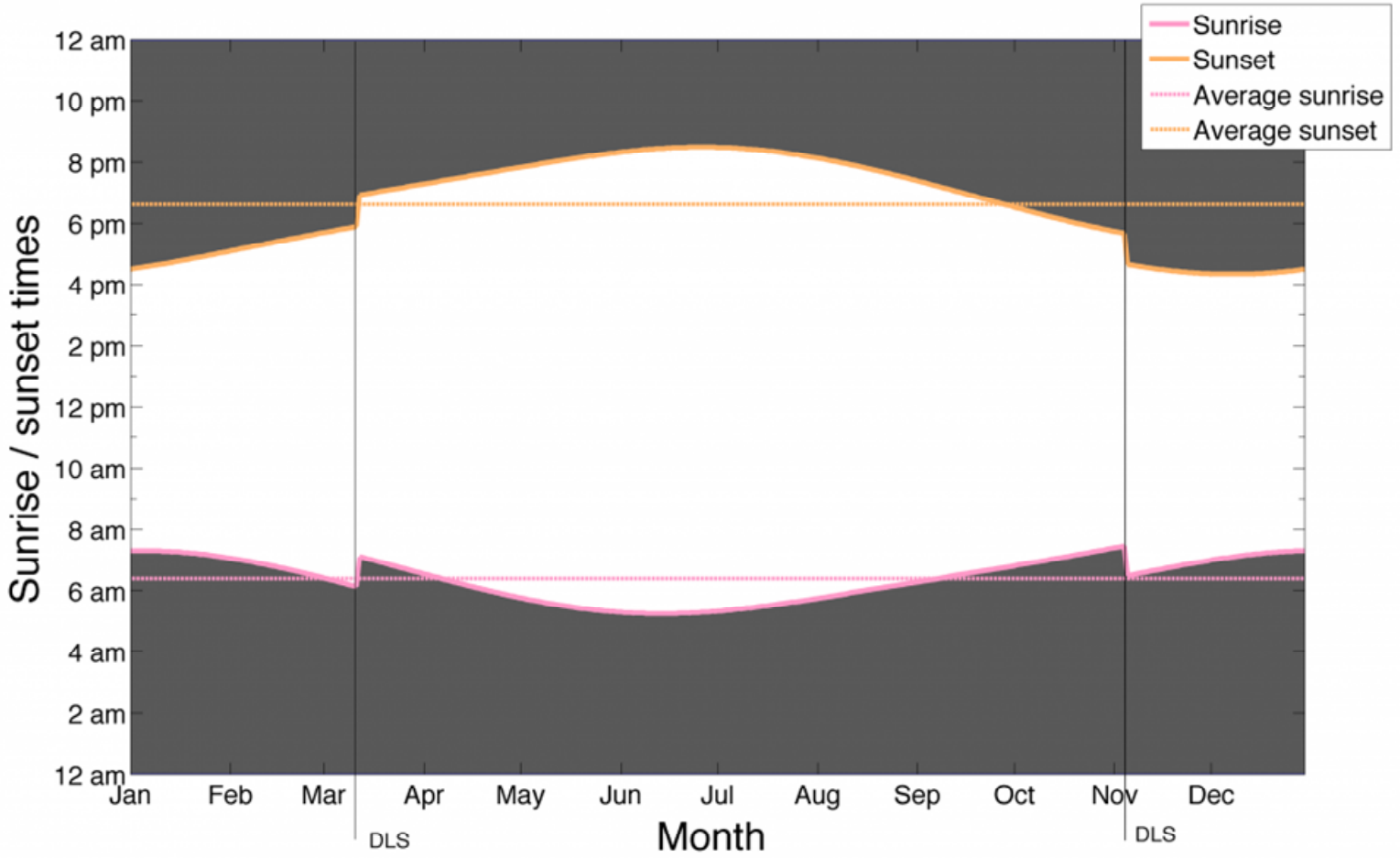
Do these tell a story?

http://www.evl.uic.edu/aej/491/week03.html



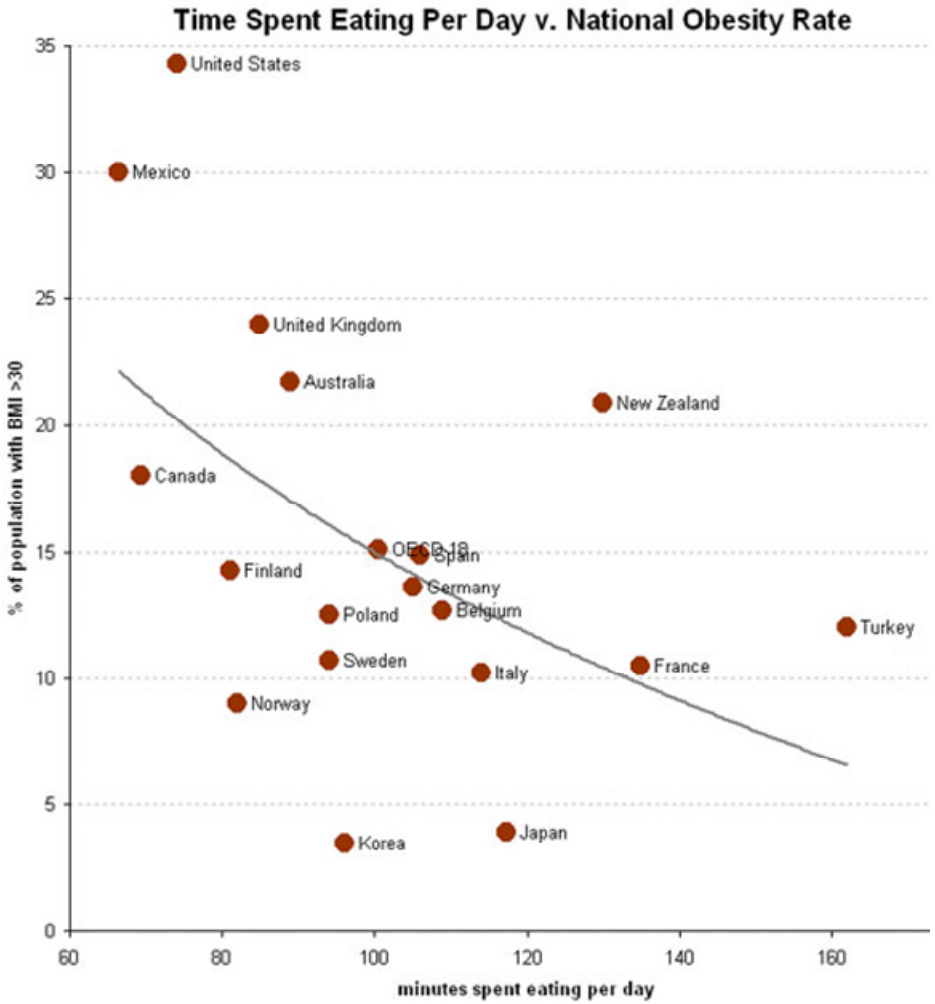
http://flowingdata.com/2009/11/26/fox-news-makes-the-best-pie-chart-ever/

Daylight Savings Time Explained



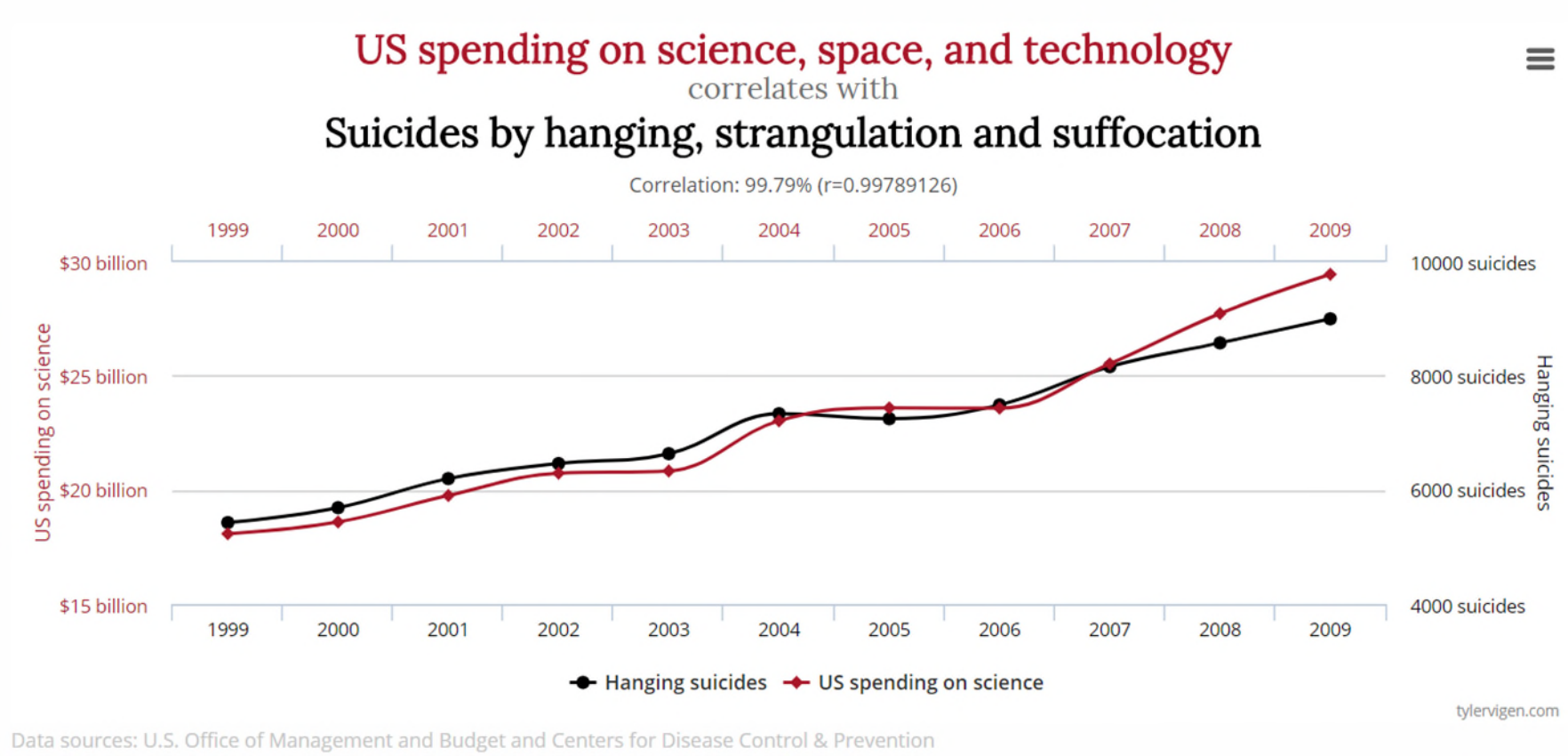
http://visual.ly/daylight-saving-time-explained

Telling a Story



http://economix.blogs.nytimes.com/2009/05/05/obesity-and-the-fastness-of-food/

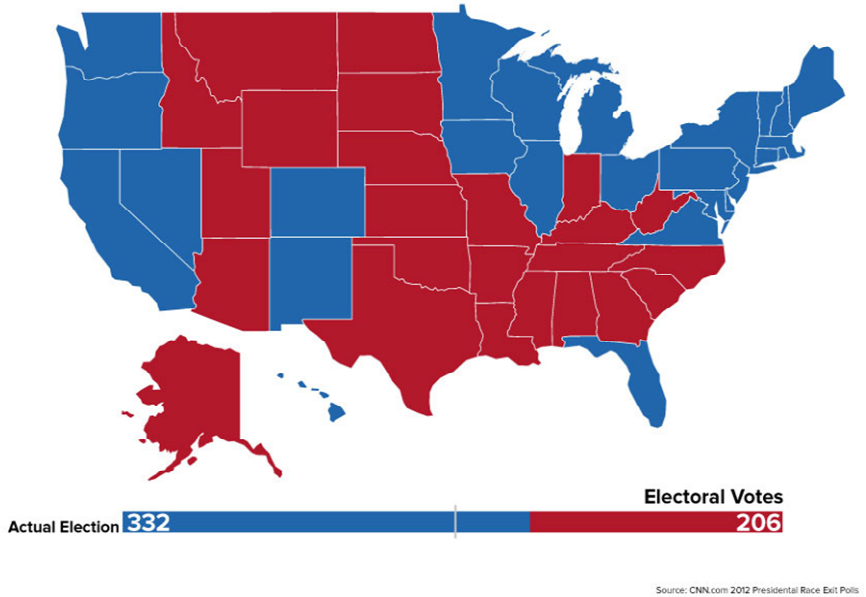
Telling a Story



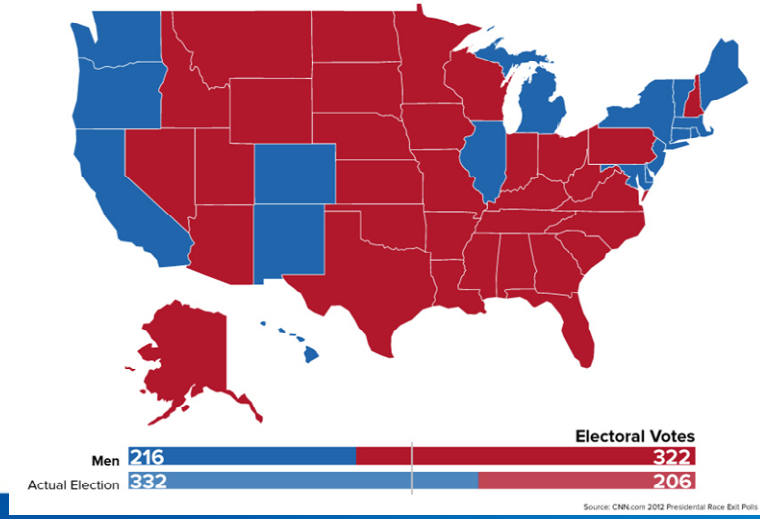
http://www.tylervigen.com/spurious-correlations

Telling a story

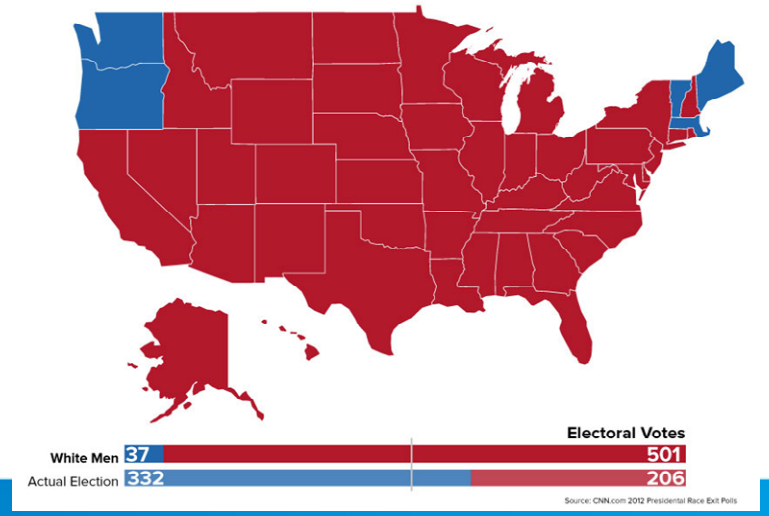
The electoral result of the 2012 election



Under pre-1920 rules: men only



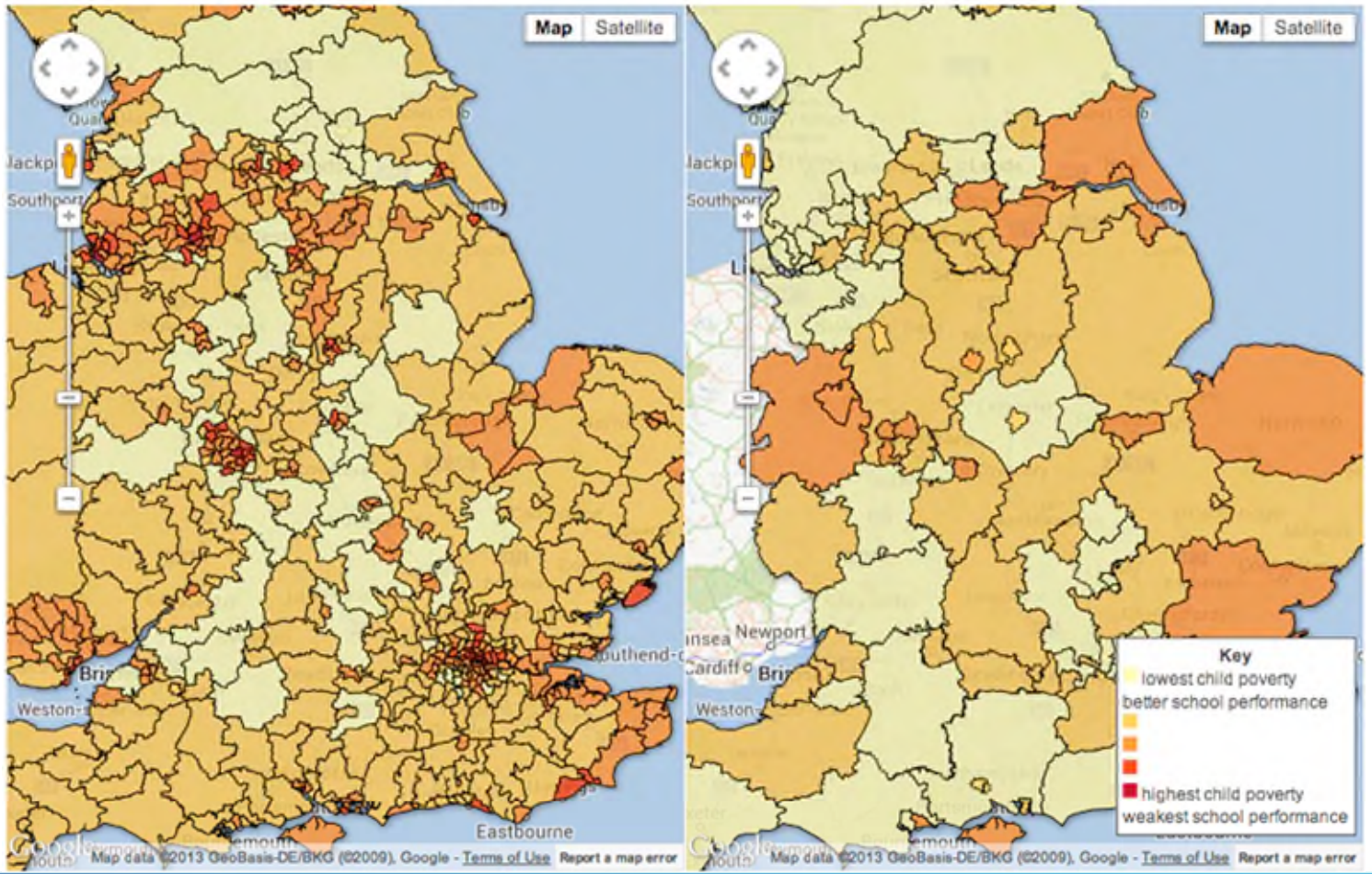
Under 1850 rules: white men only



http://www.buzzfeed.com/buzzfeedpolitics/what-the-2012-election-would-have-looked-like-with

Can this tell a story?

LEFT: Child poverty • RIGHT: Schools that 'require improvement' or are 'inadequate' according to Ofsted



Principle 2: The chart should have graphical integrity

- Basically, it shouldn't "lie" (mislead the reader)

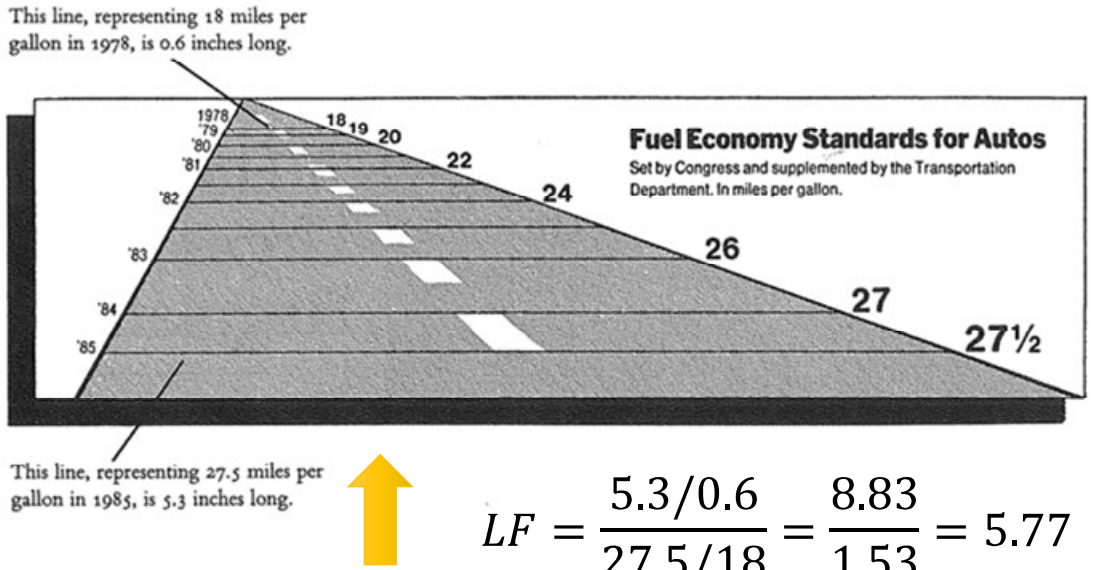
- Tufte's "Lie Factor":
  - $Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data}$

Should be ~ 1

< 1 = understated effect

> 1 = exaggerated effect

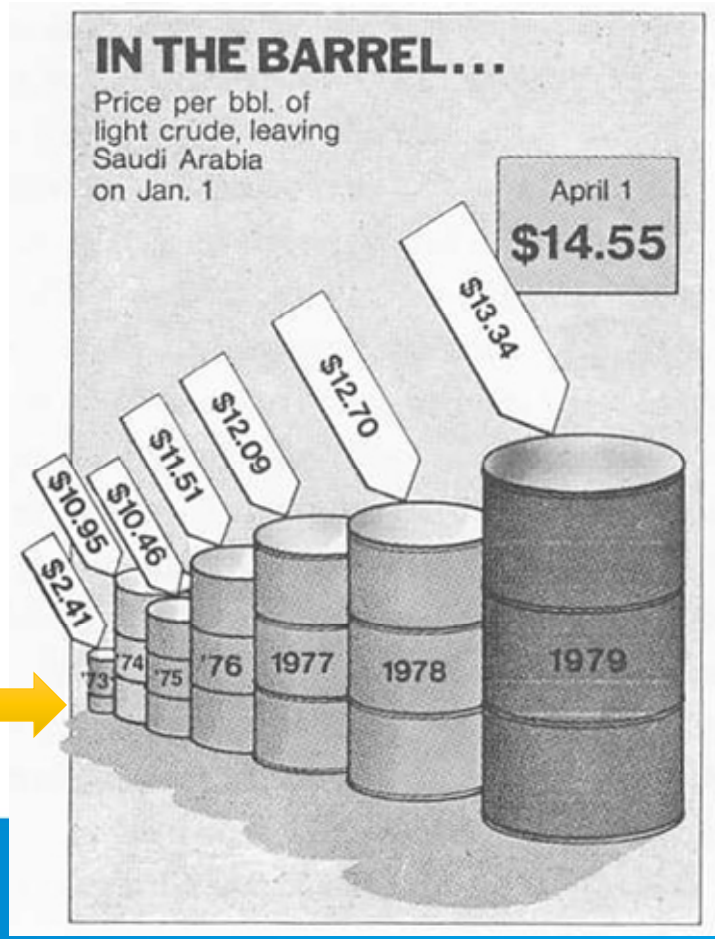
Examples of the “lie factor”



$$LF = \frac{5.3/0.6}{27.5/18} = \frac{8.83}{1.53} = 5.77$$

Reprinted from Tufte (2009), p. 57 & p. 62

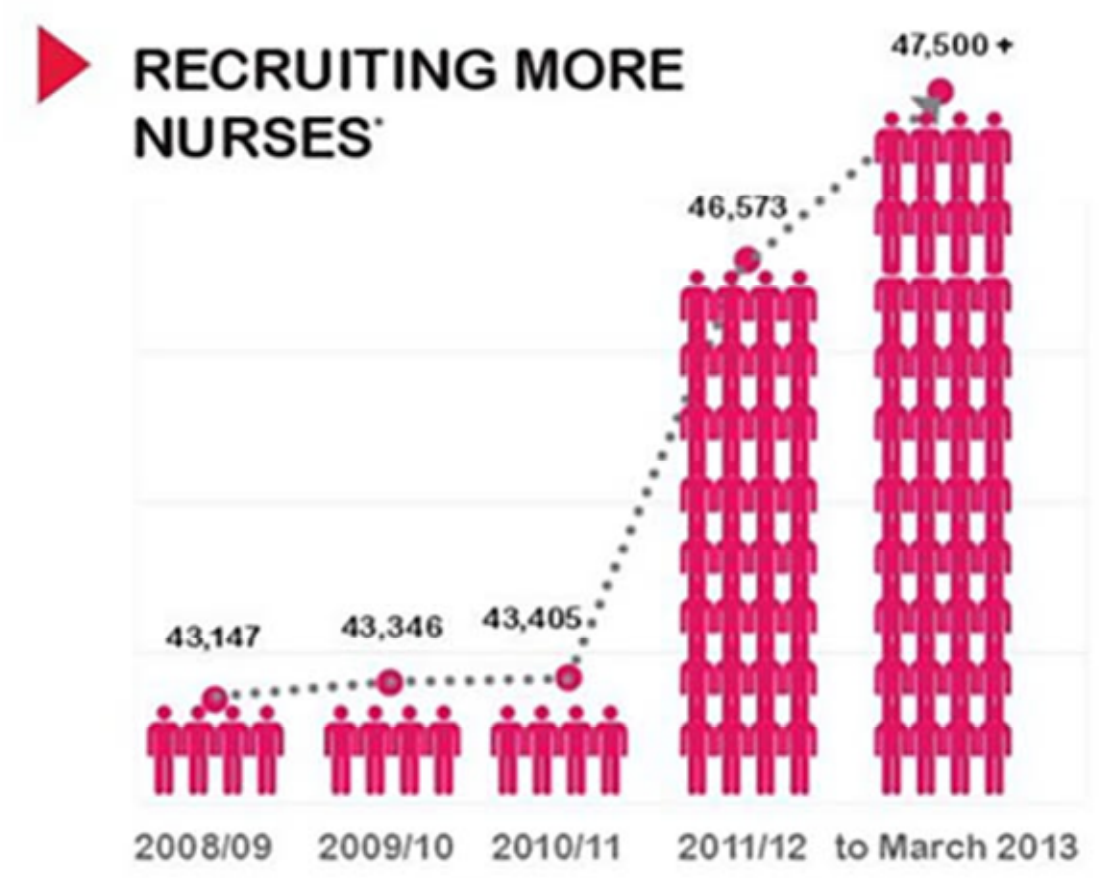
$$LF = \frac{4280\%\ (change\ in\ volume)}{454\%\ (change\ in\ price)} = 9.4$$



How is this deceptive?

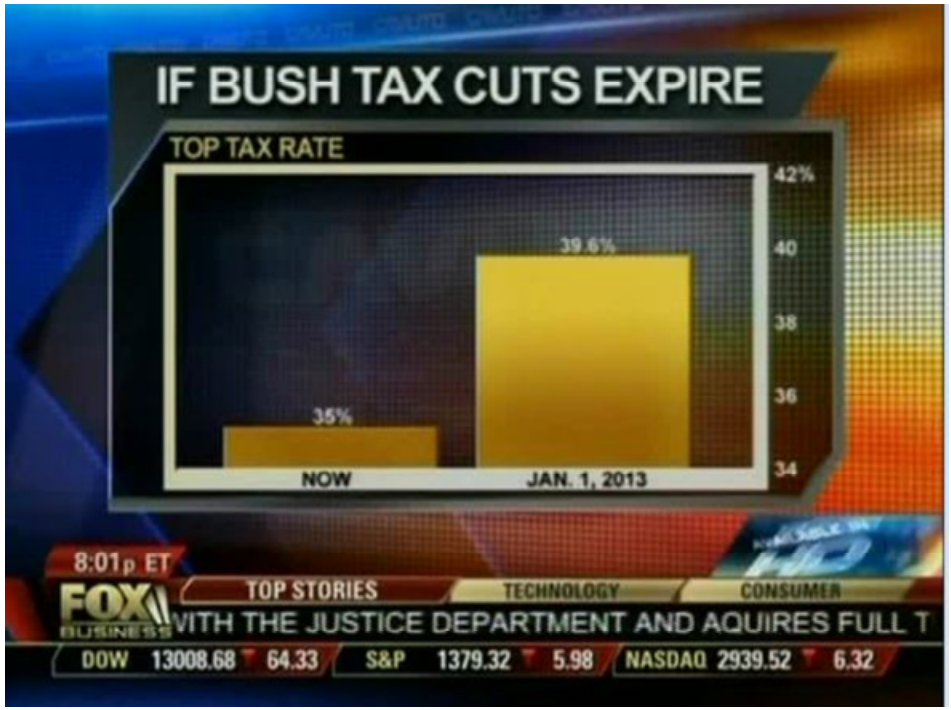
http://www.guardian.co.uk/environment/interactive/2009/mar/09/food-carbon-emissions

How about this?

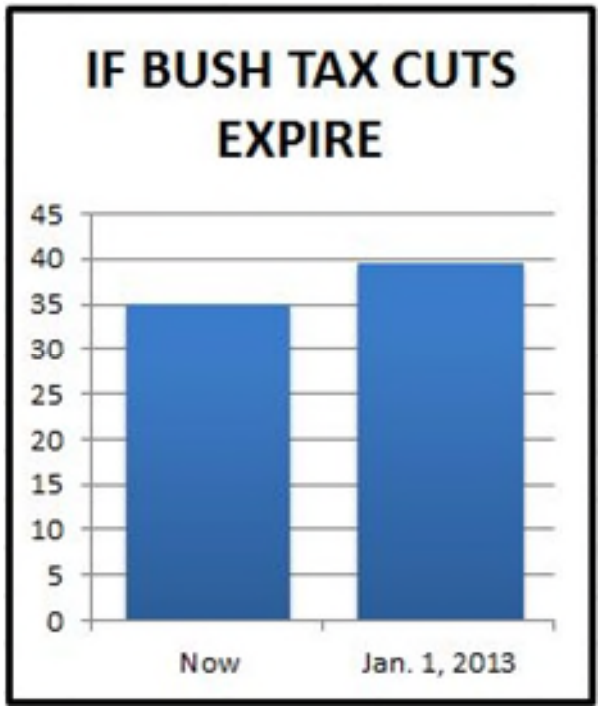


Or this?

The original graphic from Fox Business, 2012.



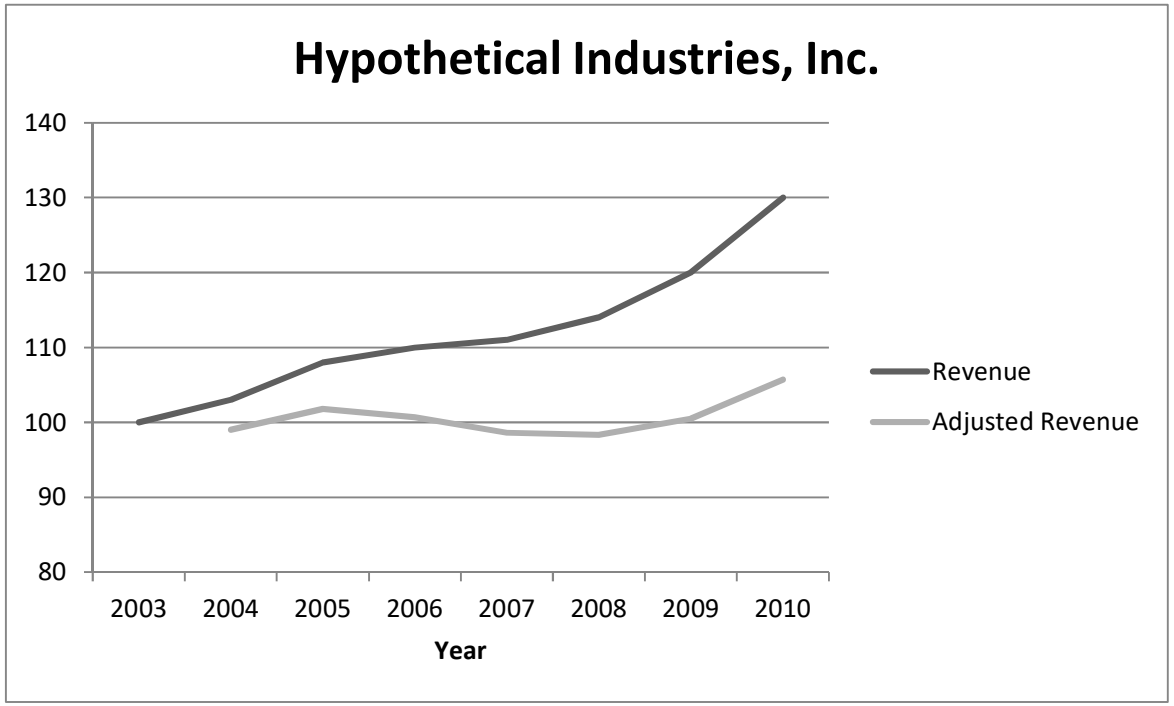
The adjusted graphic.



http://mediamatters.org/research/2012/10/01/a-history-of-dishonest-fox-charts/190225

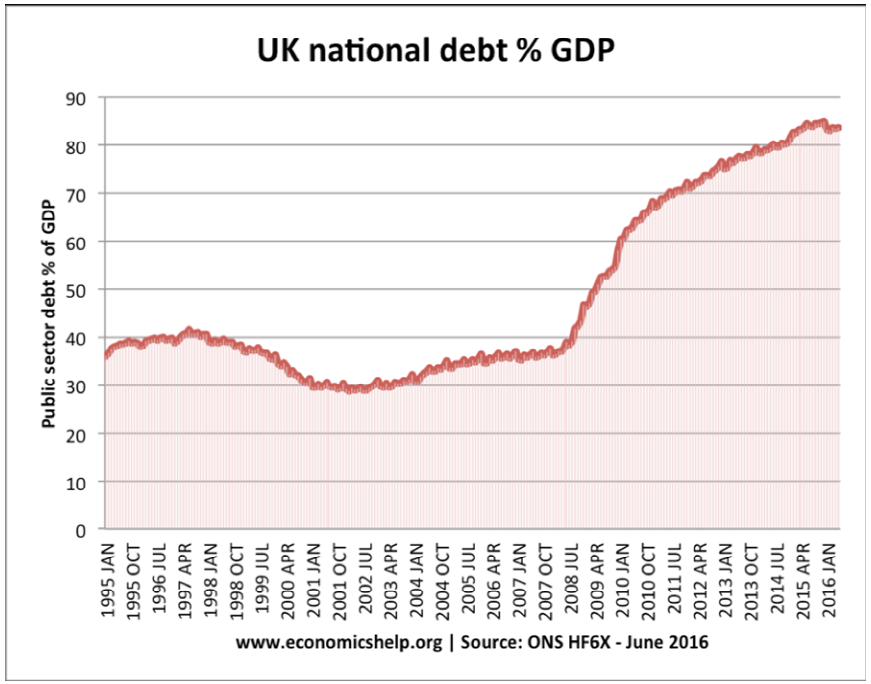
Other tips to avoid “lying”

- Adjust for inflation

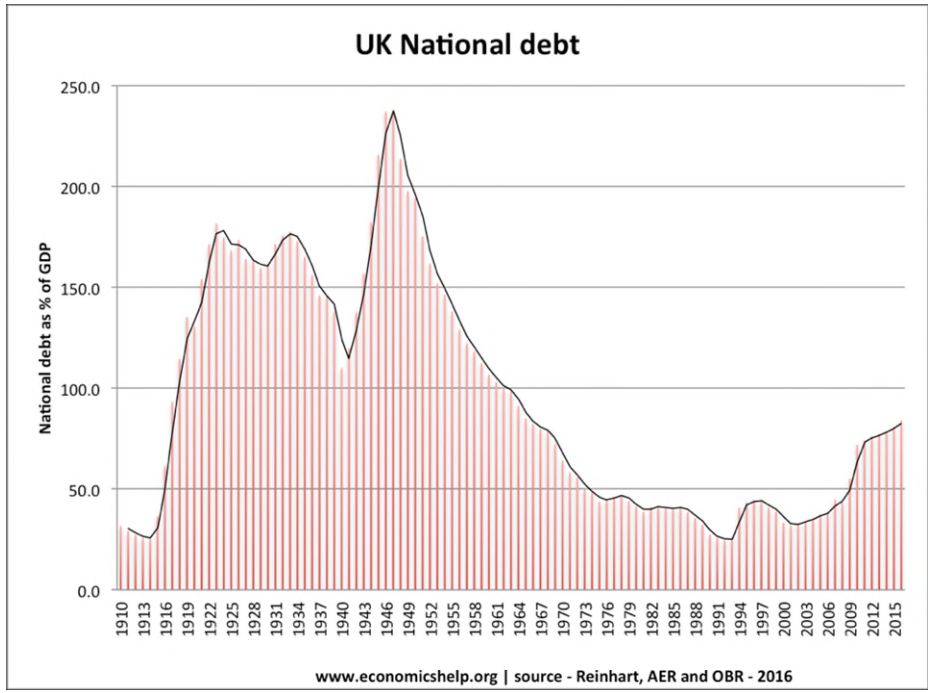


Other tips to avoid “lying”

- Make sure the context is presented



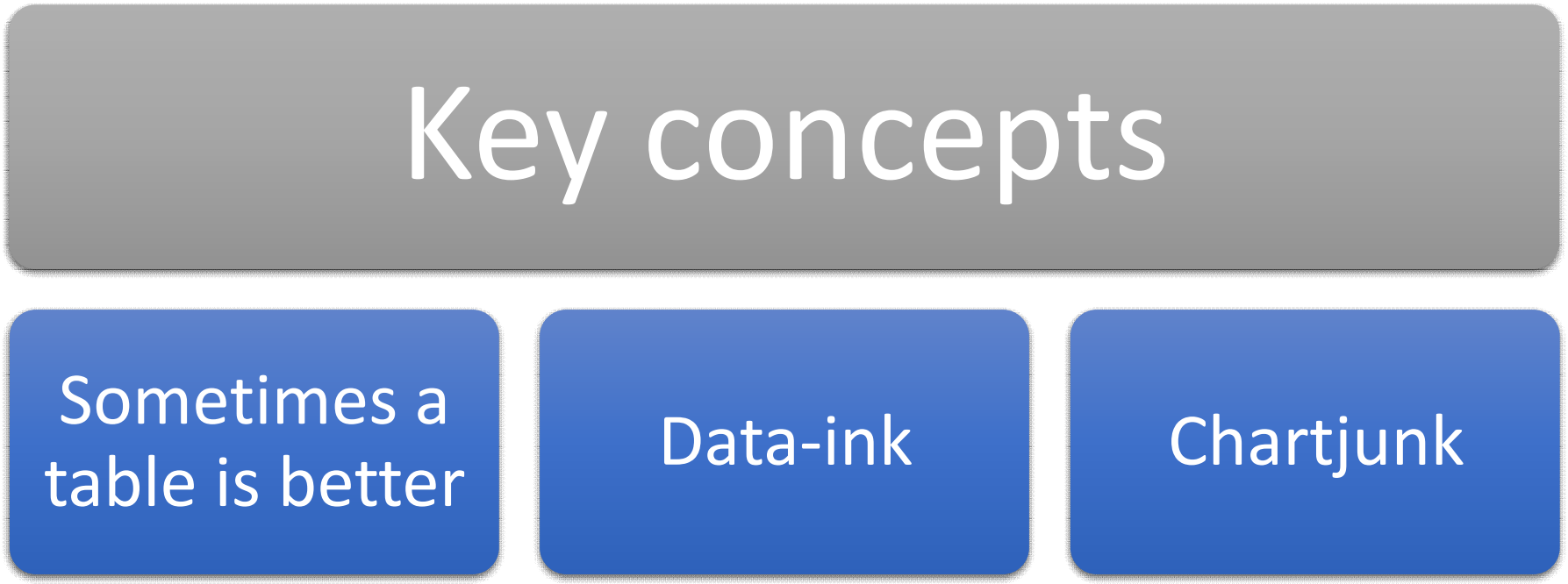
vs.



http://www.economicshelp.org/blog/21618/economics/cherry-picking-of-data/

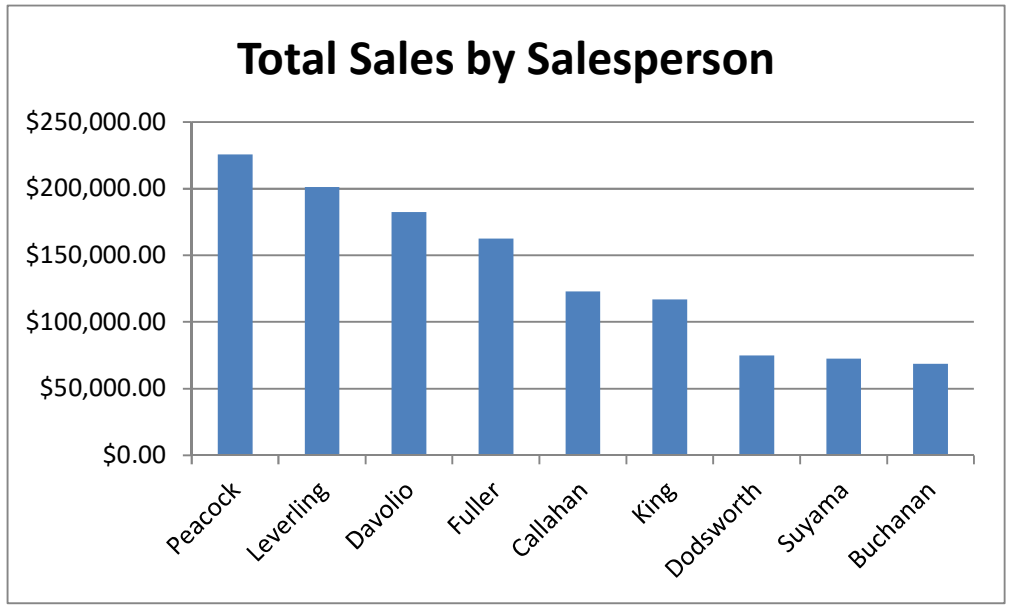
Principle 3: The chart should minimize graphical complexity

Generally, the simpler the better…



When a table is better than a chart

For a few data points, a table can do just as well…



Salesperson	Total Sales
Peacock	$225,763.68
Leverling	$201,196.27
Davolio	$182,500.09
Fuller	$162,503.78
Callahan	$123,032.67
King	$116,962.99
Dodsworth	$75,048.04
Suyama	$72,527.63
Buchanan	$68,792.25

The table carries more information in less space
and is more precise.

The Ultimate Table: The Box Score

- Large amount of information in a very small space

- So why does this work?
  - Depends on the reader's knowledge of the data

R	Team	Tournament	Total	OutOfBox	SixYardBox	PenaltyArea	Rating
1	Paris Saint Germain	Ligue 1	17.3	5.5	1.2	10.6	7.24
2	Barcelona	La Liga	15.4	5	1.4	9	7.18
3	Manchester City	Premier League	17.5	6.6	1.4	9.5	7.17
4	Bayern Munich	Bundesliga	18.7	6.9	1.4	10.3	7.15
5	Real Madrid	La Liga	18.3	5.7	1.4	11.3	7.10
6	Juventus	Serie A	15.2	6.9	0.9	7.4	7.09
7	Atletico Madrid	La Liga	10.9	4.6	0.6	5.6	7.02
8	Monaco	Ligue 1	13.1	4.8	1	7.3	7.00
9	Manchester United	Premier League	13.8	5.4	1.2	7.2	7.00
10	Liverpool	Premier League	17.3	6.9	1.3	9.1	7.00
11	Marseille	Ligue 1	16.4	6	1.7	8.7	6.99
12	Napoli	Serie A	17.5	7.5	1.2	8.9	6.98
13	Tottenham	Premier League	17.2	7.3	1.1	8.7	6.97
14	Lyon	Ligue 1	14.4	6.5	0.7	7.3	6.97
15	Roma	Serie A	17.7	7.3	1.2	9.2	6.97
16	Lazio	Serie A	13.9	5.4	1	7.5	6.96
17	Chelsea	Premier League	16.2	6.8	0.8	8.5	6.95
18	Borussia Dortmund	Bundesliga	14.2	4.3	1.3	8.5	6.93
19	Valencia	La Liga	12.2	5	0.9	6.3	6.91
20	Arsenal	Premier League	15.8	6.1	1	8.7	6.90

© WhoScored
Page 1/5 | Showing 1 - 20 of 98 first | prev | next | last

Data Ink

- The amount of “ink” devoted to data in a chart
- Tufte’s Data-Ink ratio:
  - $Data - ink\ ratio = \frac{data-ink}{total\ ink\ used\ in\ graphic}$

Should be ~ 1

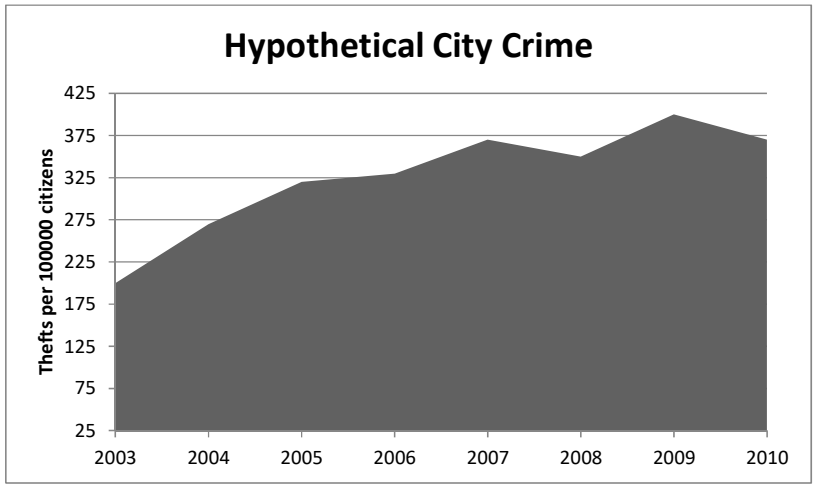
< 1 = more non-data related ink in graphic

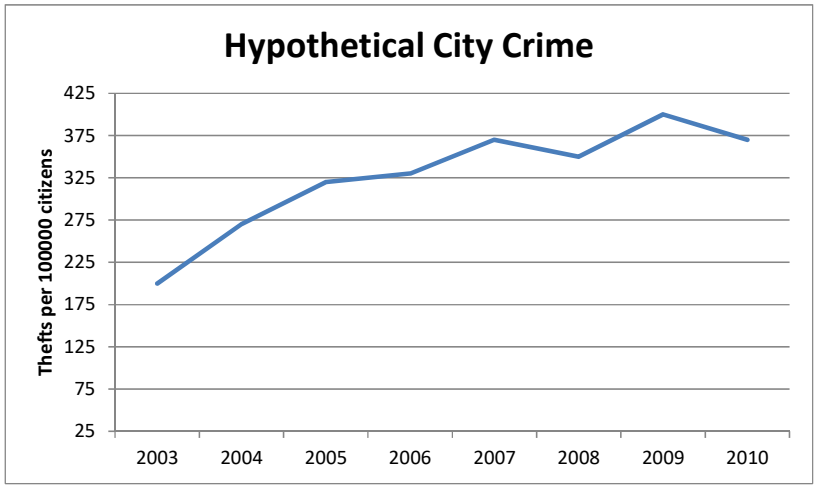
= 1 implies all ink devoted to data

Tufte’s principle:
Erase ink whenever possible

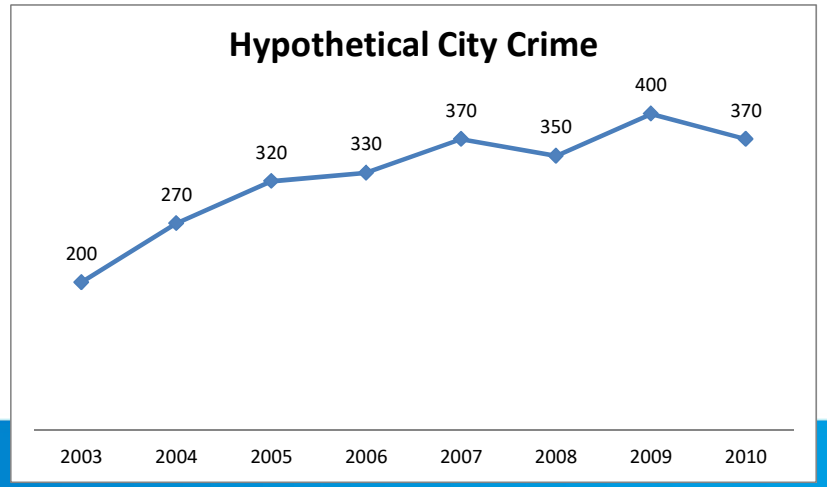
Being conscious of data ink

Lower data-ink ratio (worse)

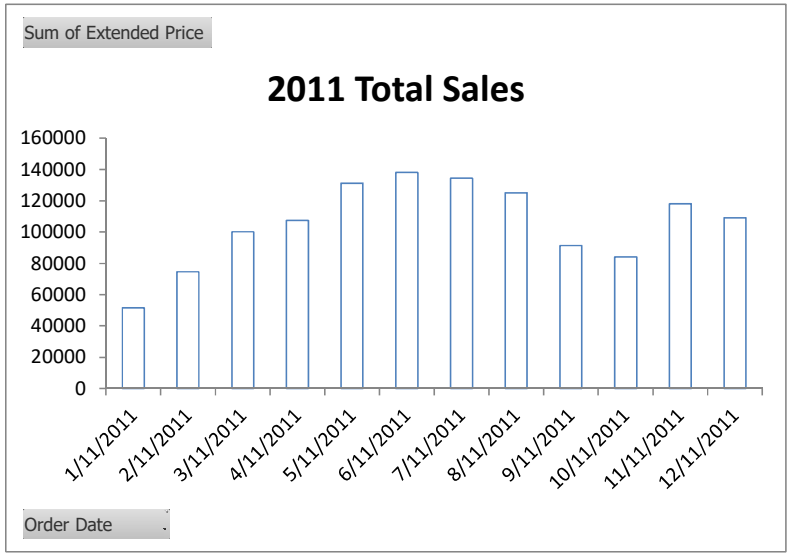


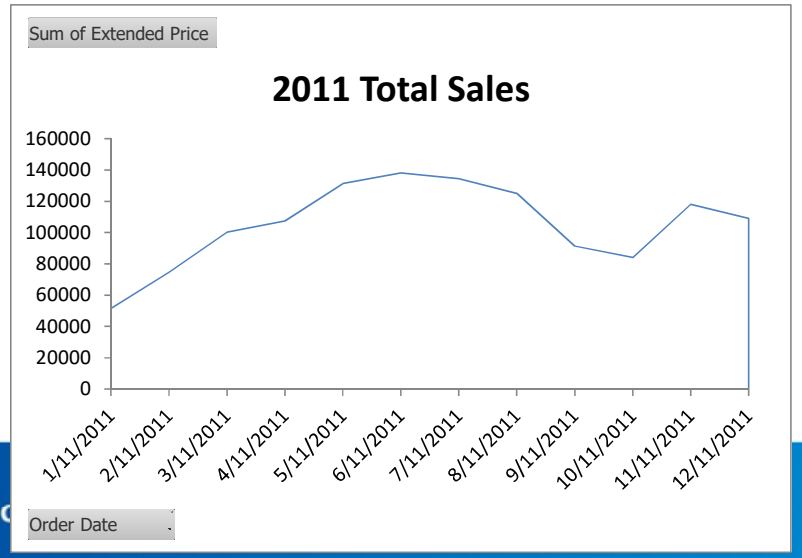


Higher data-ink ratio (better)



What makes a good chart?



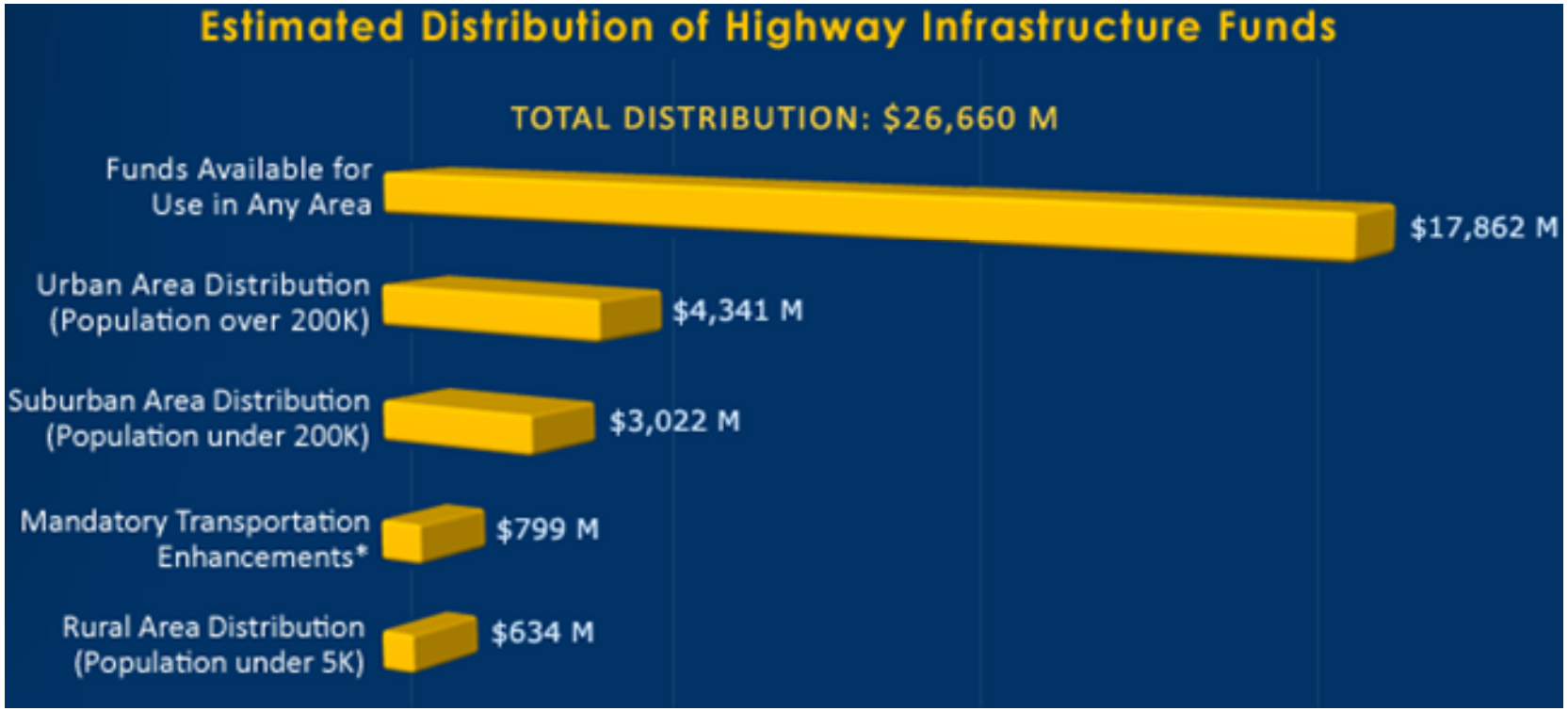


Sometimes it's really a matter of preference.

These both minimize data ink.

Why isn't a table better here?

3-D Charts



Evaluate this from a data-ink perspective.
How does it affect the clarity of the chart?

Source: www.Recovery.gov (website has been taken down)

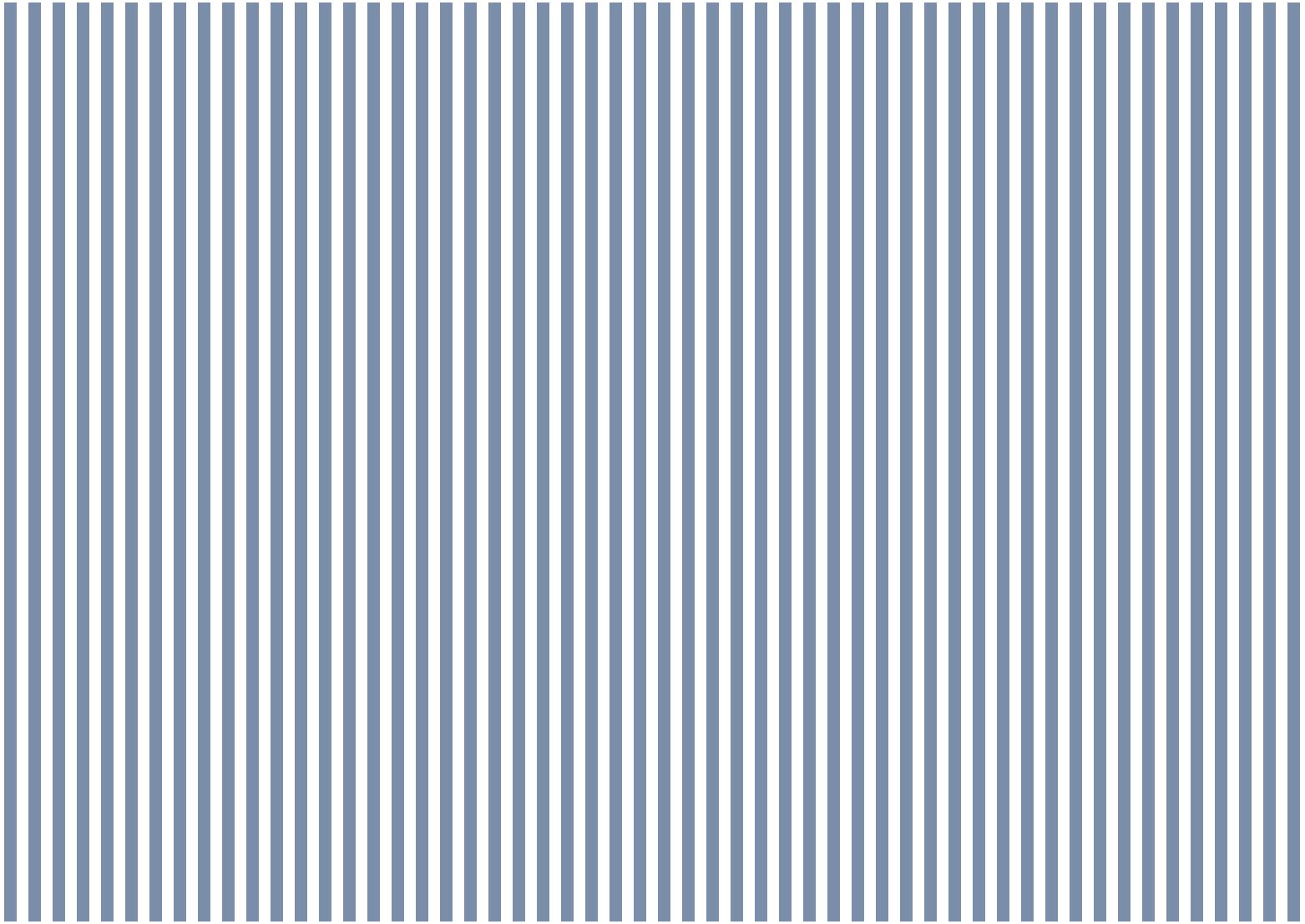
Chartjunk: Data Ink “gone wild”

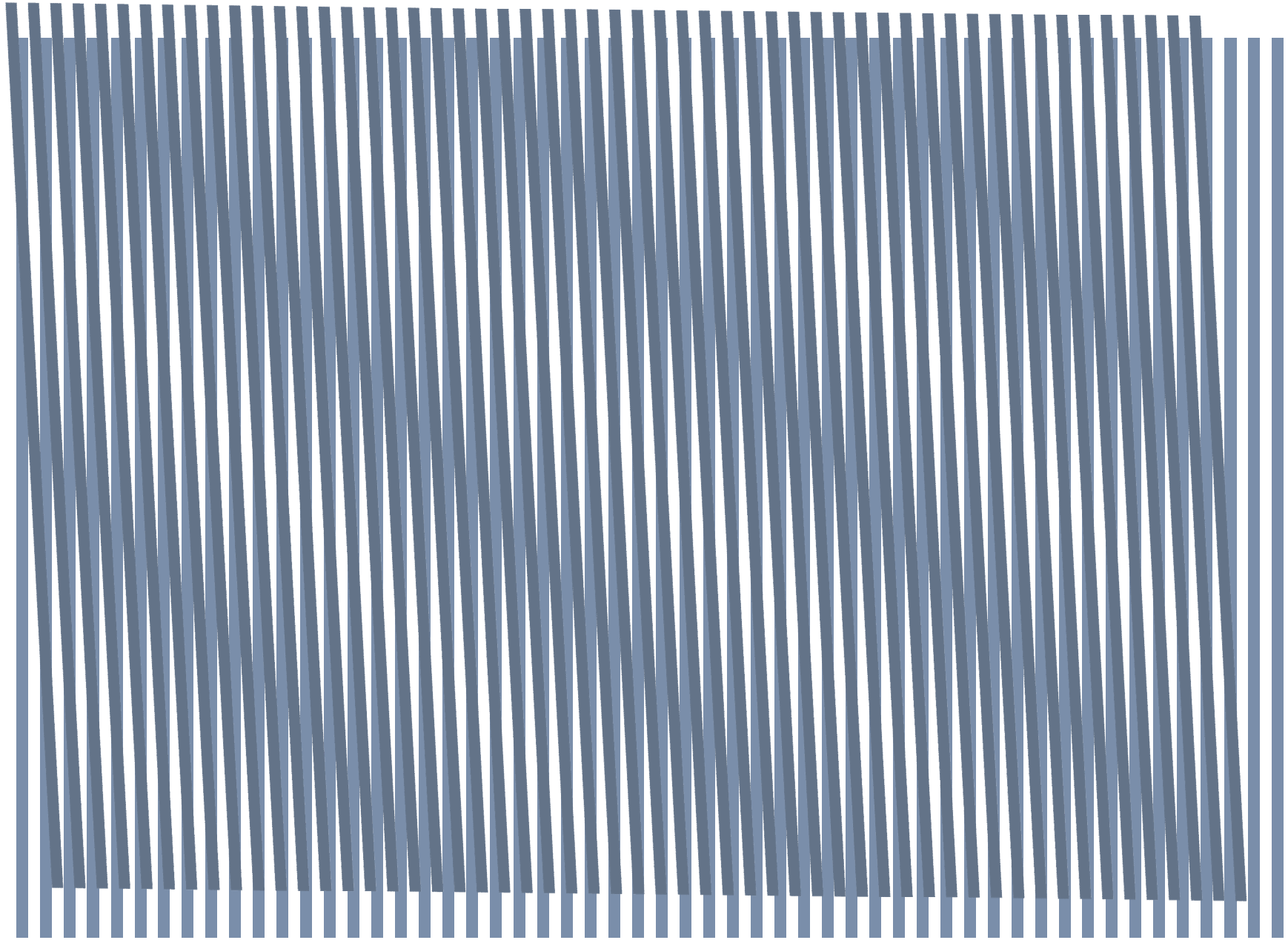
Unnecessary visual clutter that doesn’t provide additional insight

Distraction from the story the chart is supposed to convey

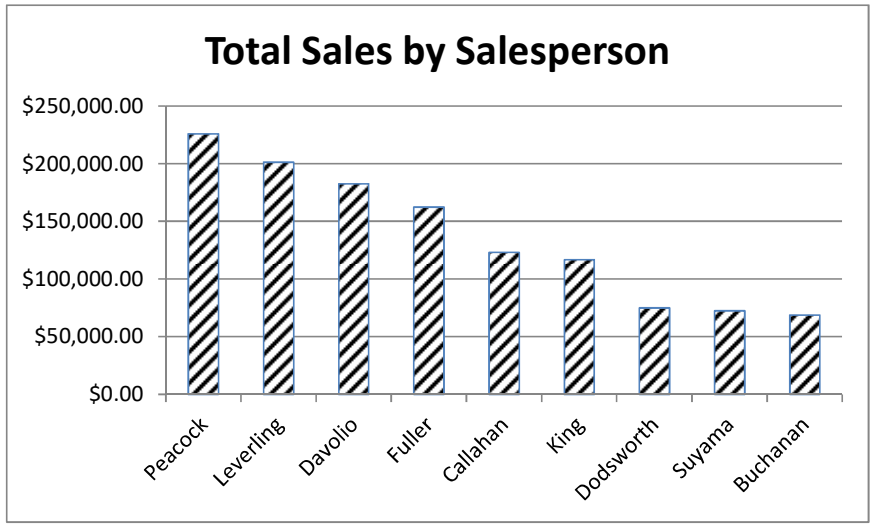
When the data-ink ratio is low, chartjunk is likely to be high

The Moiré effect

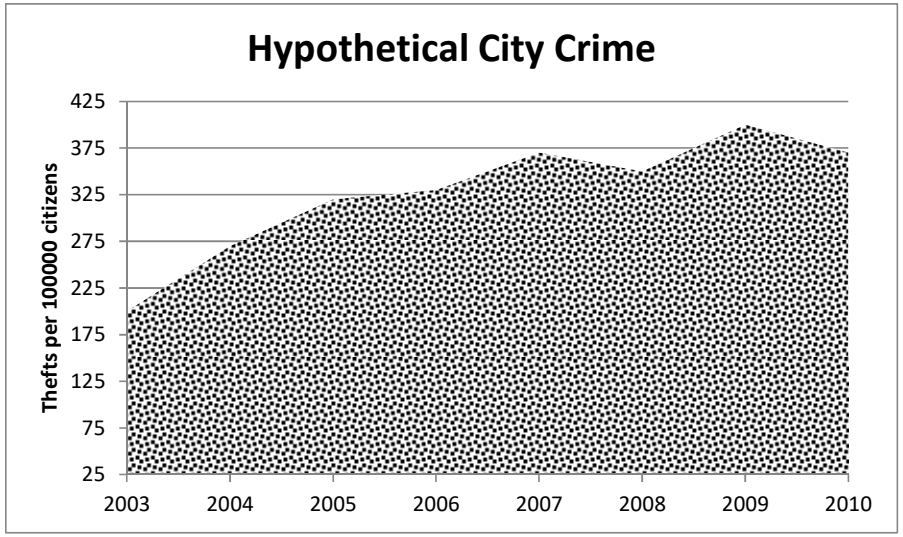




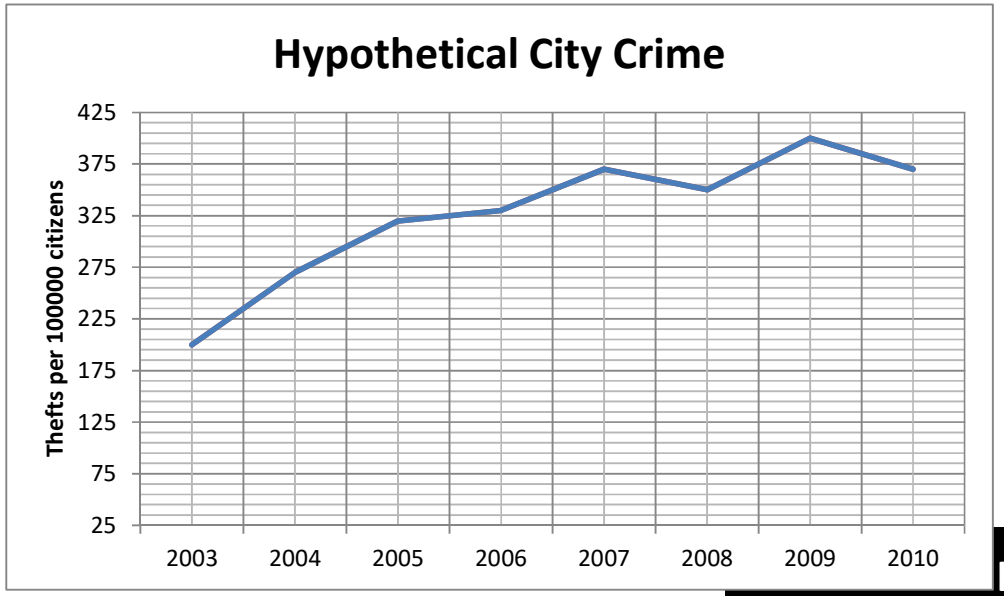
Example: Moiré effects (Tufte 2009)



Stands out, in a bad way

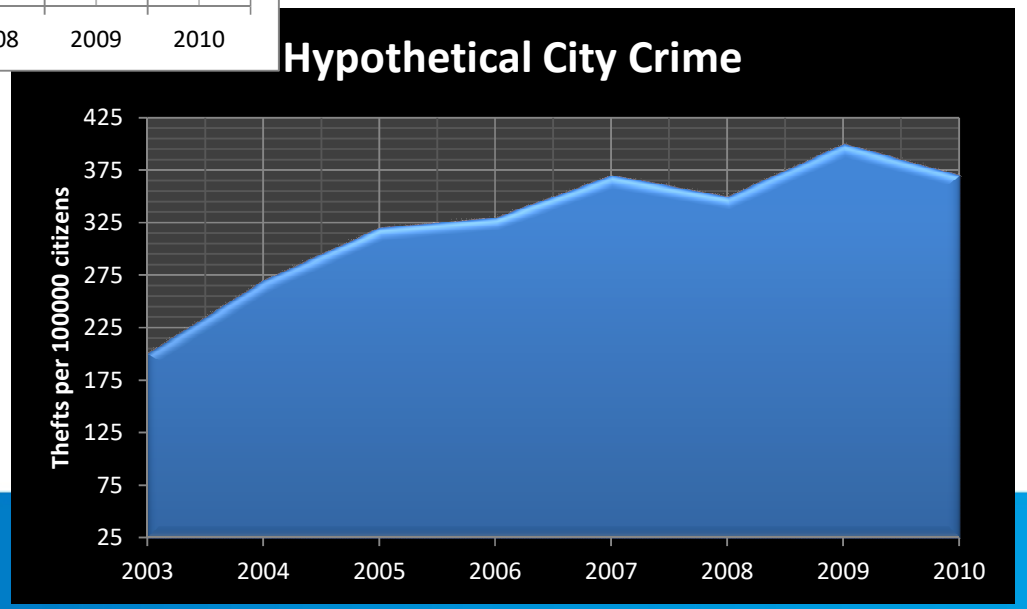


Example: The Grid

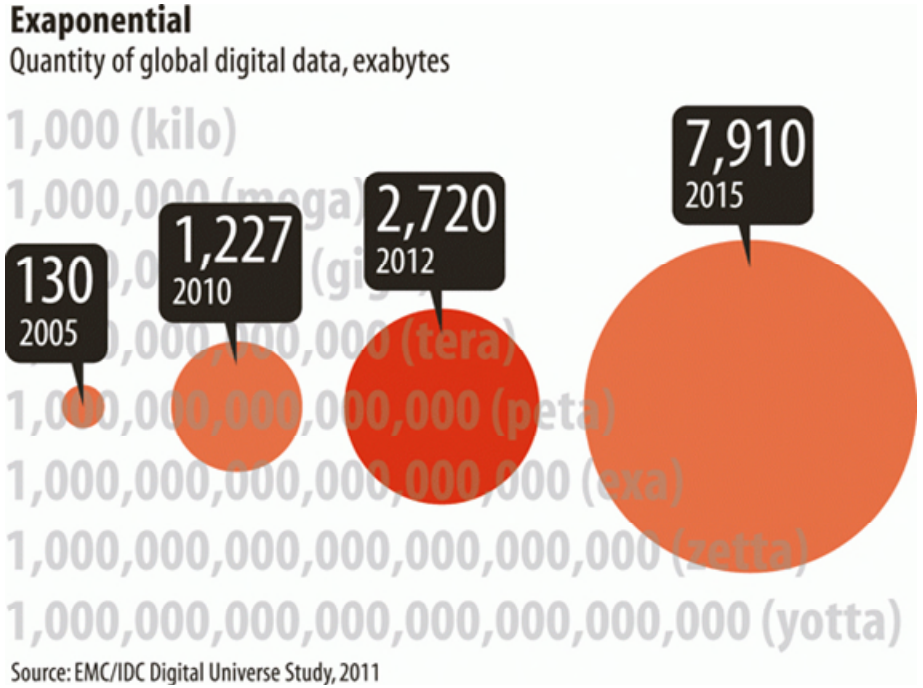


Why are these examples of chartjunk?

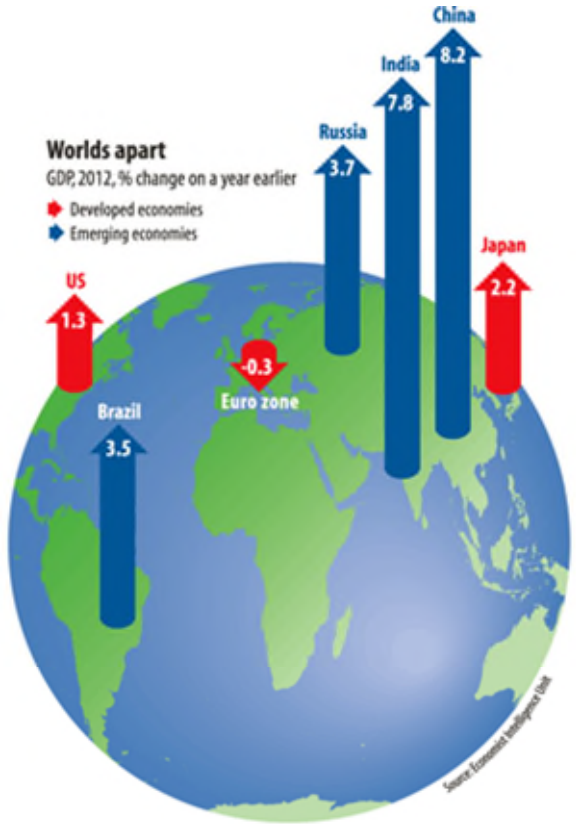
What could you do to remedy it?



Data Ink Working Against Us

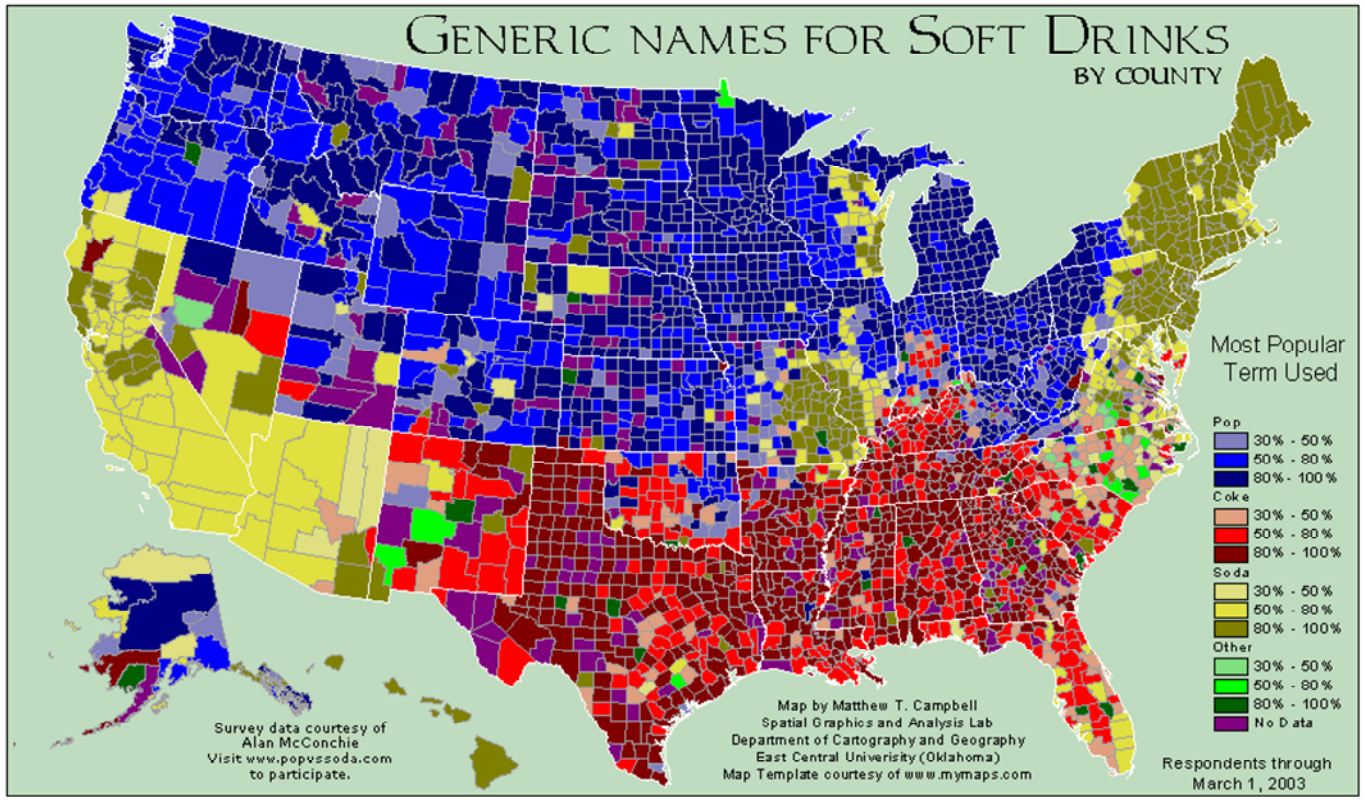


http://www.economist.com/node/21537922



http://www.economist.com/node/21537909

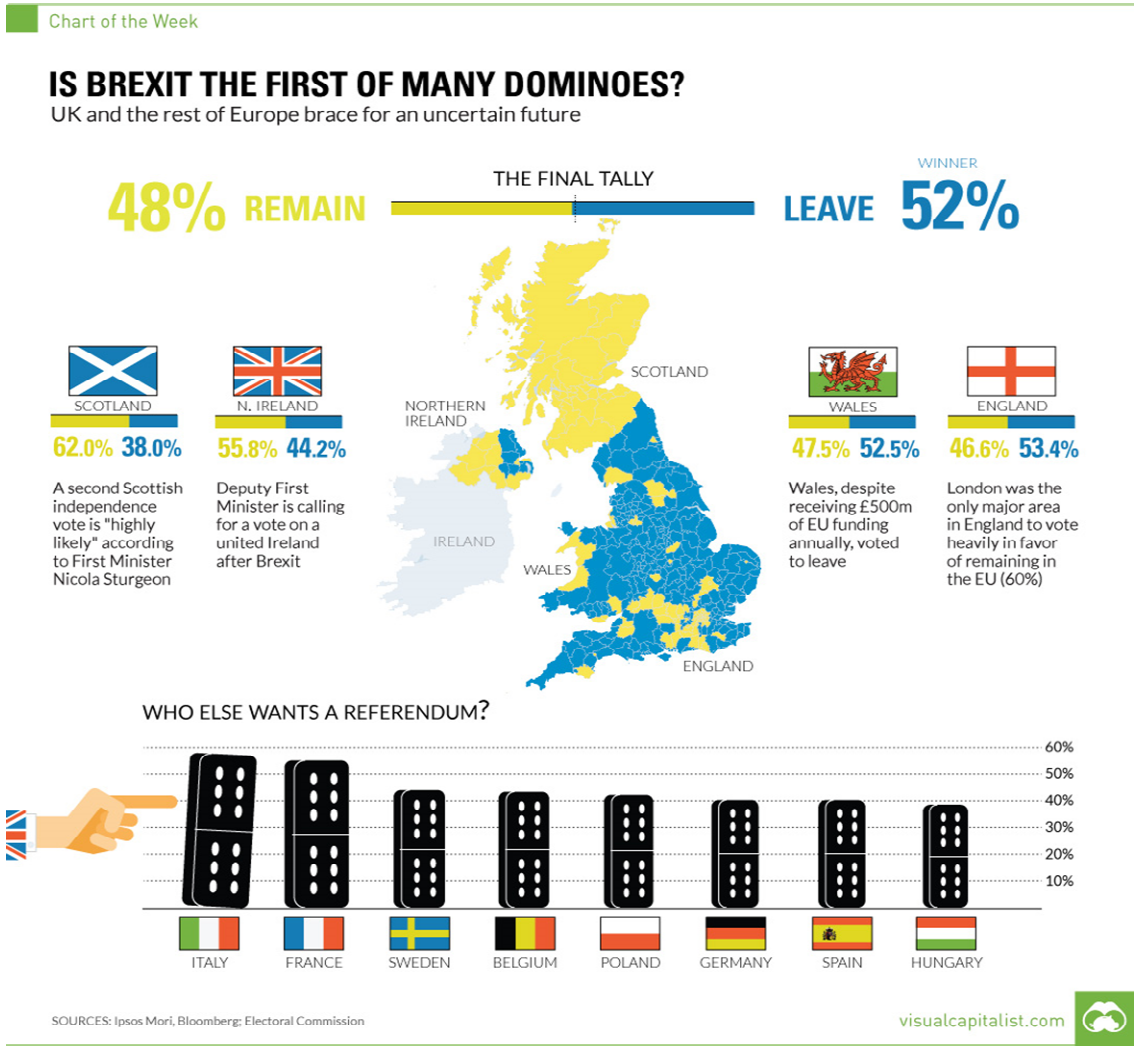
Data Ink Working For Us



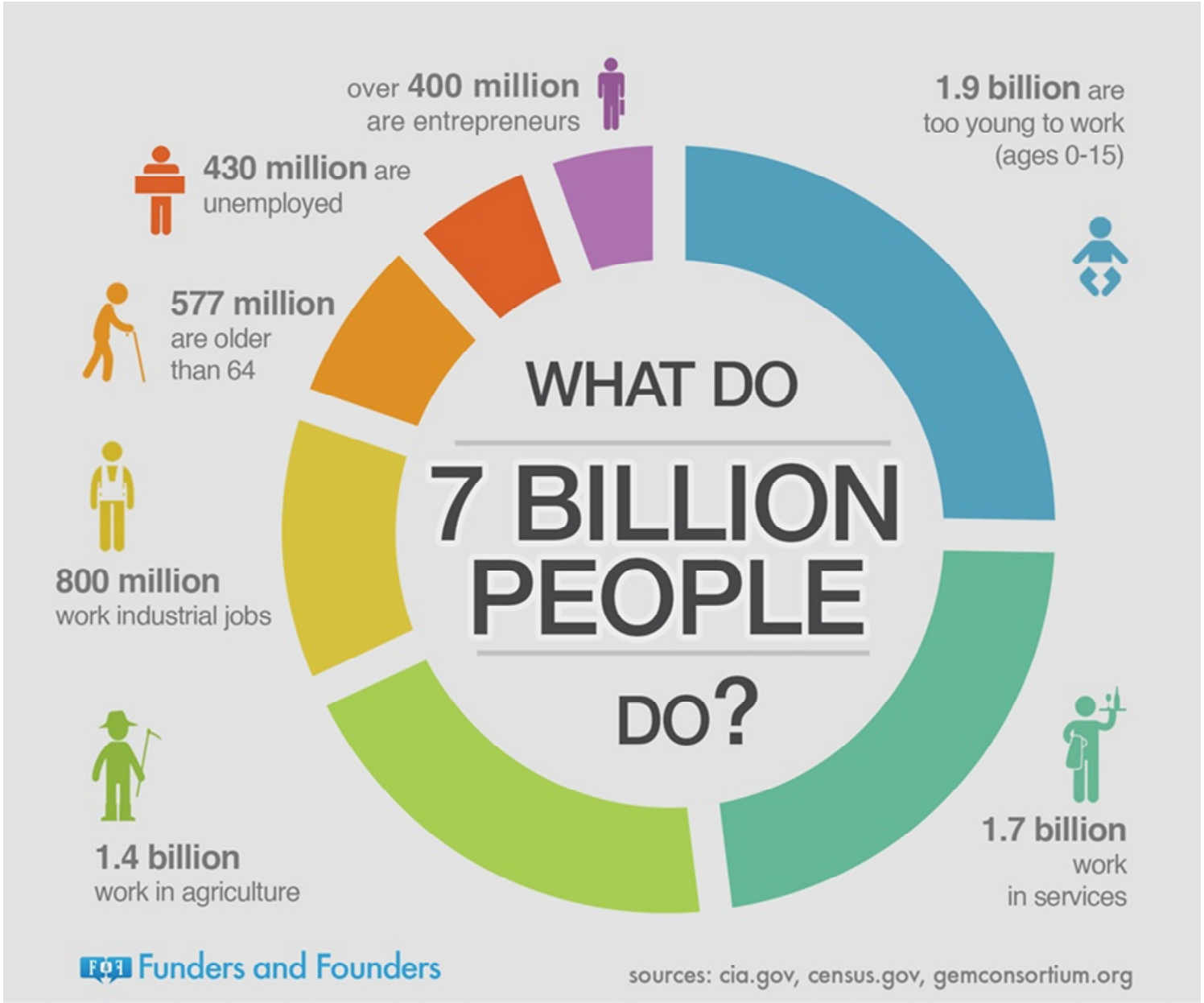
Evaluate this chart in terms of Data Ink.

Imagine this as a bar chart. As a table!!

Three excerpts from an infographic on Brexit



http://www.visualcapitalist.com/brexit-first-many-dominoes/


over 400 million are entrepreneurs
1.9 billion are too young to work (ages 0-15)
430 million are unemployed
577 million are older than 64
WHAT DO
7 BILLION PEOPLE
DO?
800 million work industrial jobs
1.4 billion work in agriculture
1.7 billion work in services
Funders and Founders
sources: cia.gov, census.gov, gemconsortium.org

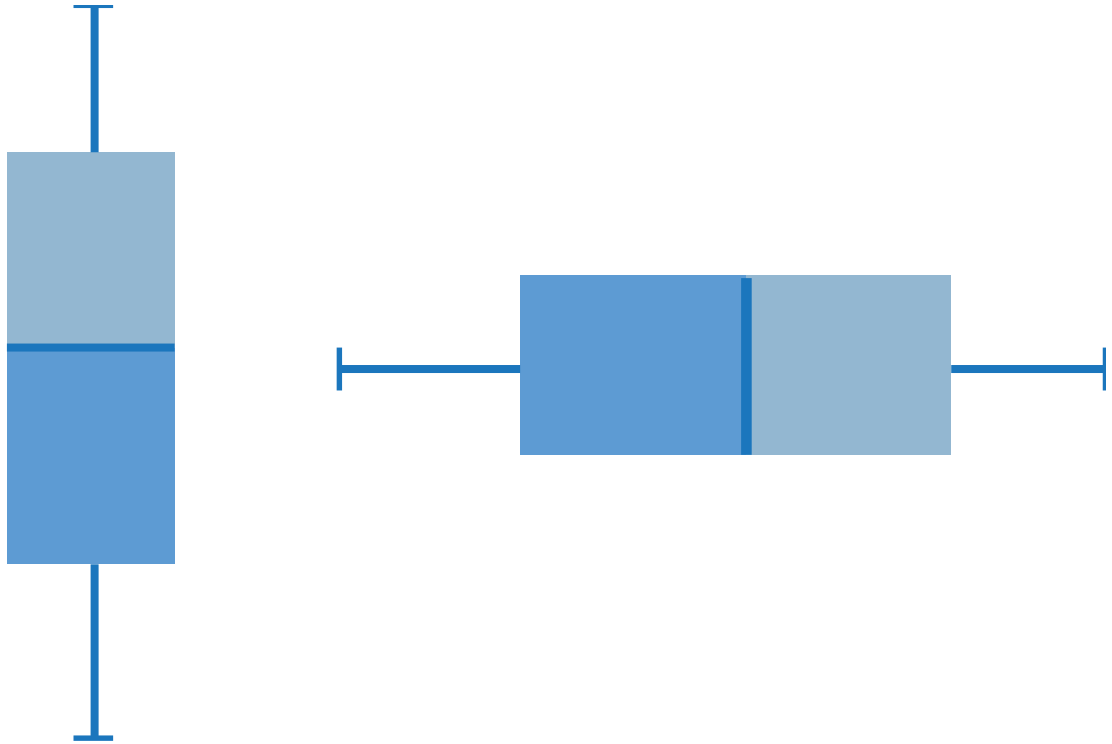
Visualization with Tableau

Bar Chart

encodes data using height/length of bar and shows categorical comparisons.

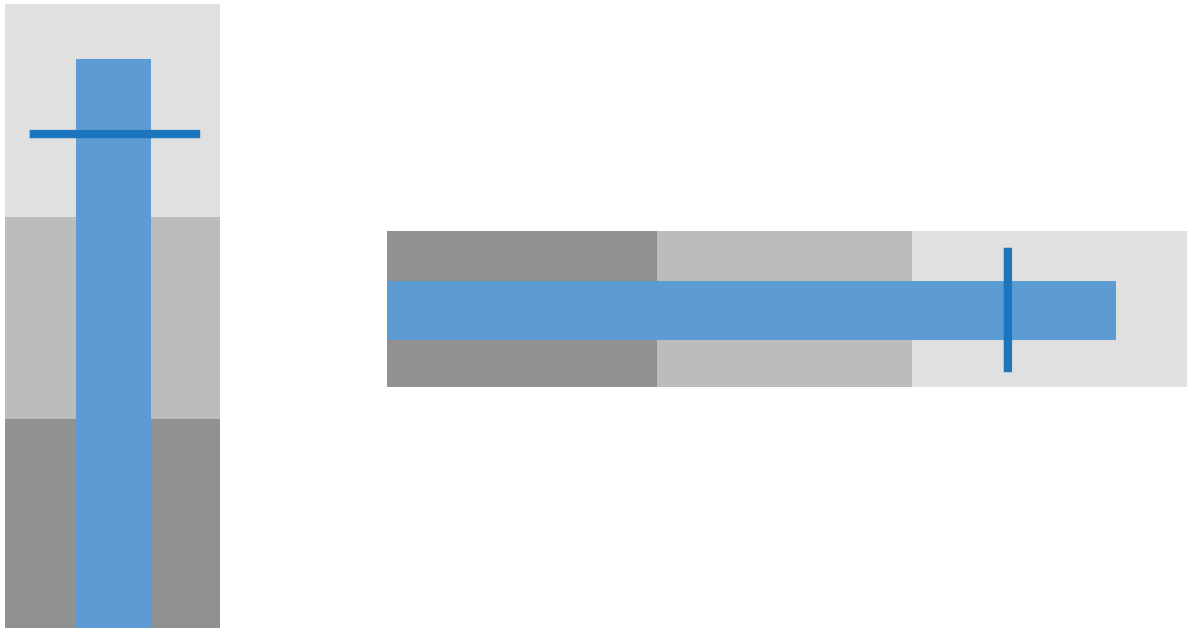
Box Plot

encodes data using position and height/length to show the distribution of the data.



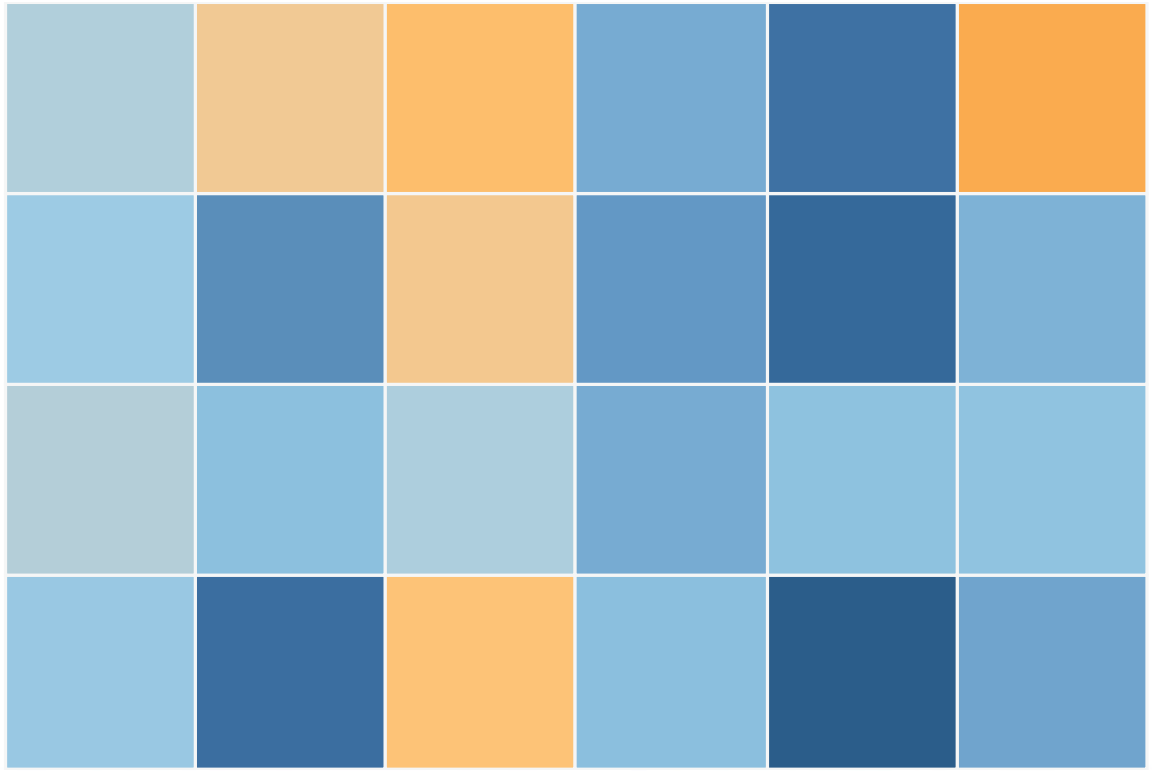
Bullet Graph

encodes data using length/height, position and color to show actual compared to target and performance bands (compare two measurements).



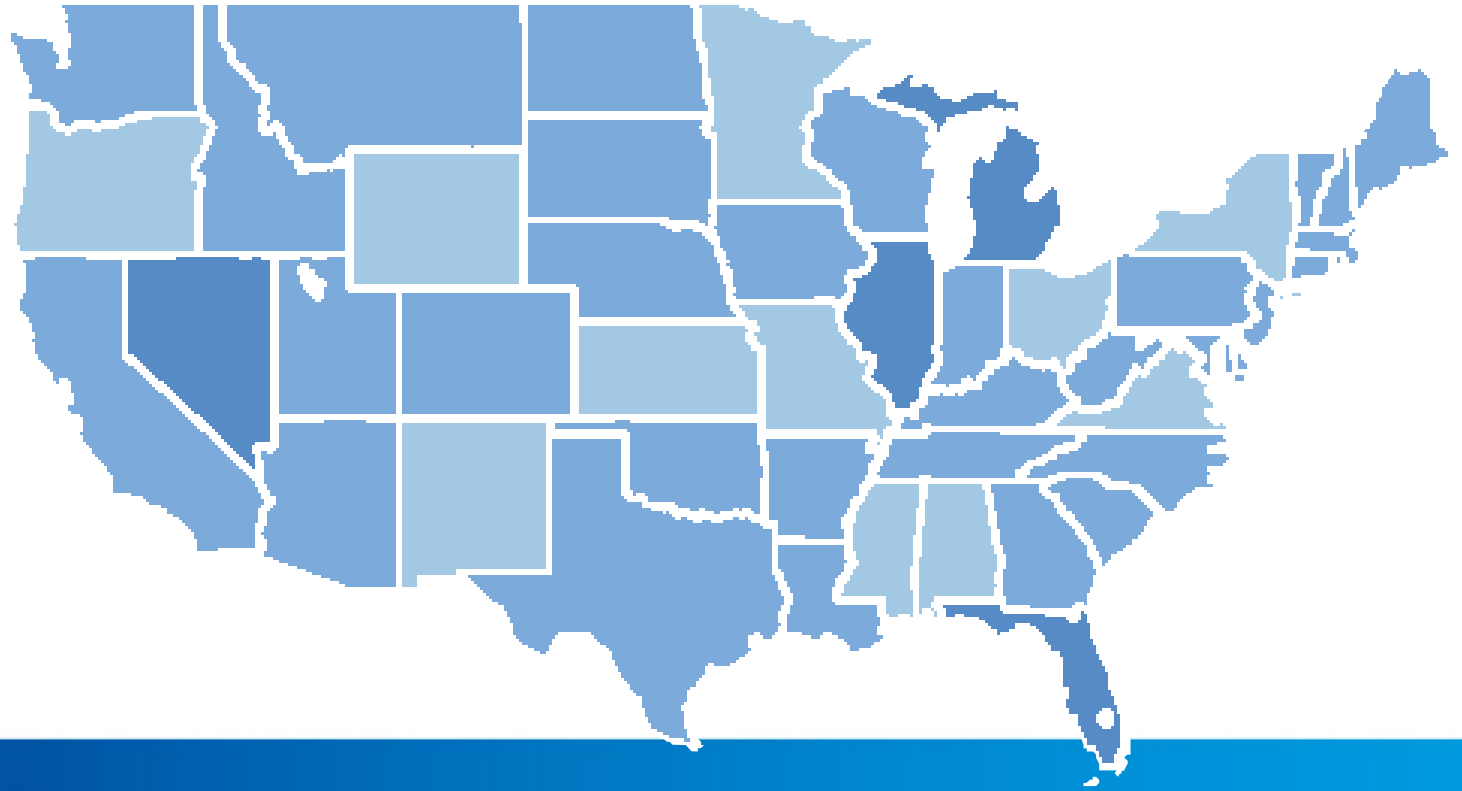
Heat Map

encodes a data table using color to highlight the differences in the table without numbers.



Choropleth Map (Shaded Map)

encodes data using color and position to show data geographically.



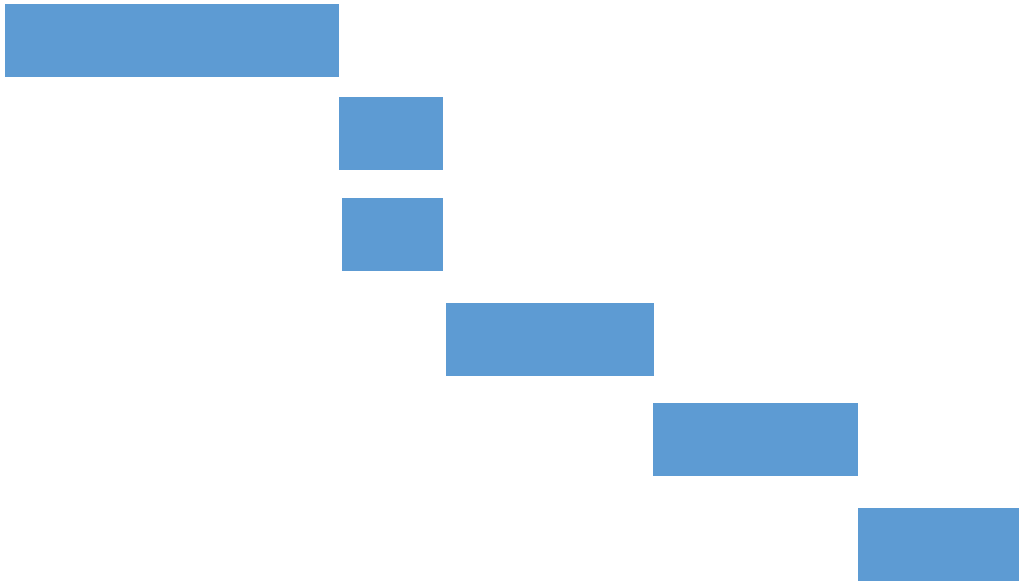
Highlight Table

encodes a data table using color to highlight the differences in the table numbers.

$29,071	$17,307	$30,073
$2,603	$2,353	$5,079
$66,106	$53,891	$42,444
$20,173	$14,151	$26,664
$100,615	$58,304	$98,684
$71,613	$35,768	$70,533
$10,760	$8,319	$18,127
$39,140	$43,916	$84,755

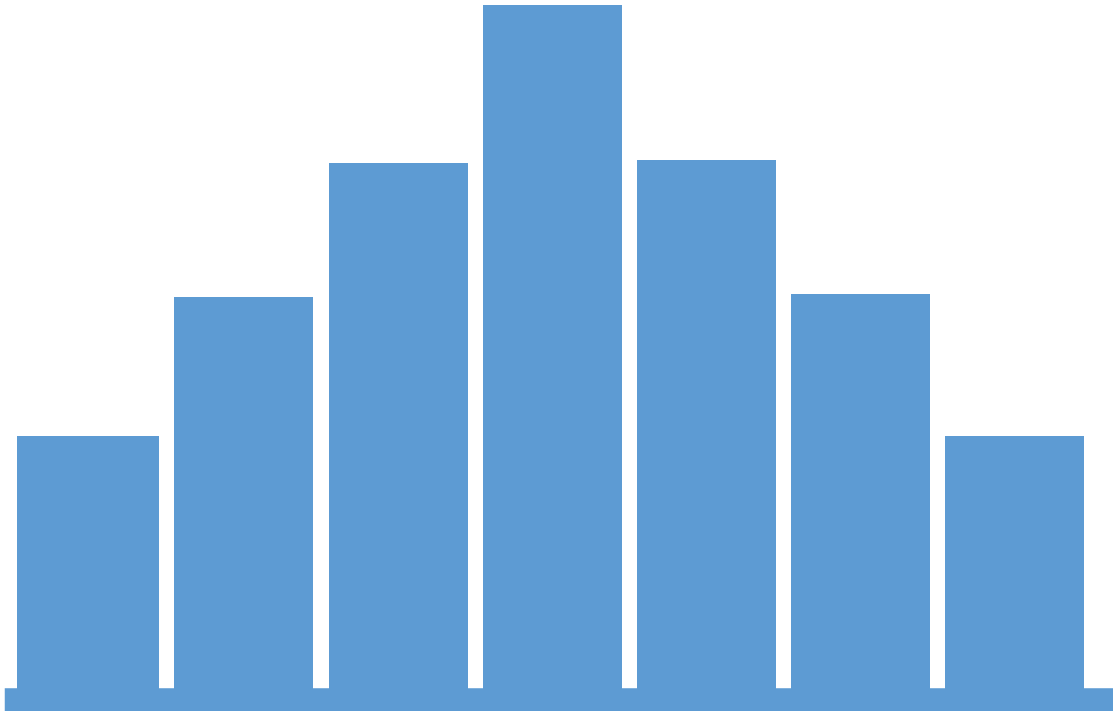
Gantt Chart

encodes data using length and position to show amount of work completed in segments of time.



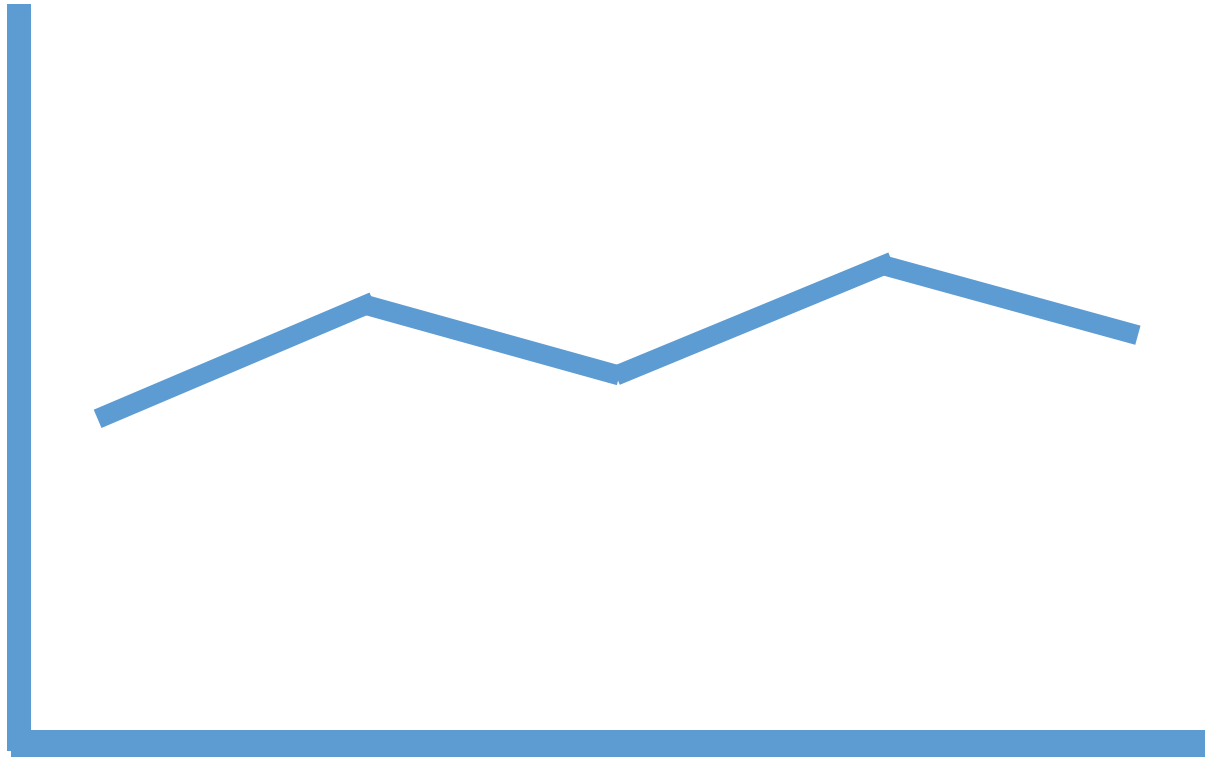
Histogram

encodes data using height and shows a distribution.



Line Chart

encodes data using position and often shows trend over time.



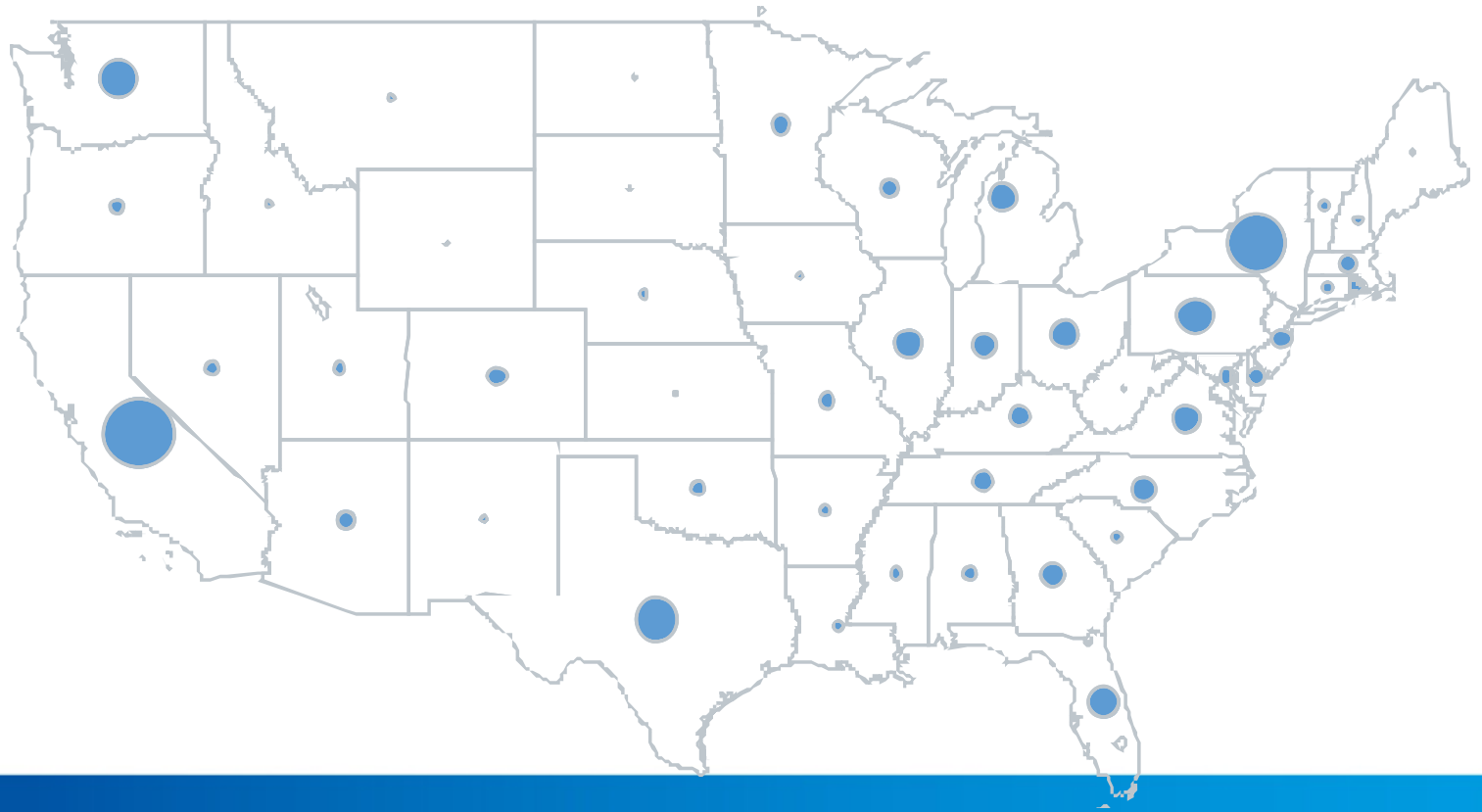
Stacked Bar Chart

encodes data using height or length of bar and color by segment and shows categorical and part-to-whole comparisons.

* Caution be careful not to slice stacked charts into too many segments.

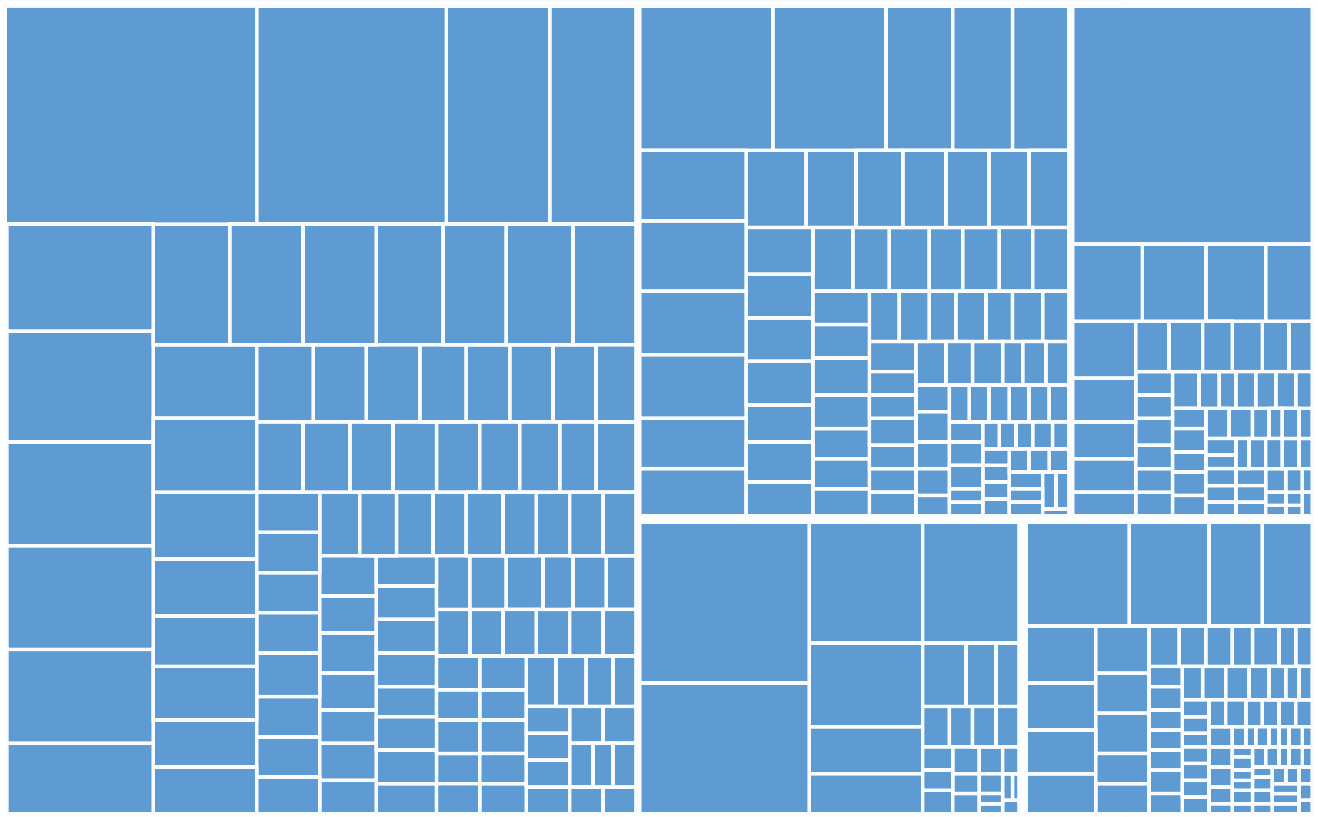
Symbol Map (Dot Map)

encodes data using position to show data geographically and can also use size to show quantitative data.



Treemap

encodes data using size and color and is useful for hierarchical data or when there are a very large number of categories to compare.



Bubble Chart

encodes data using size of circle to show comparisons which is difficult for making precise quantitative comparisons.

* Caution this chart type is not recommended.

Working with external data

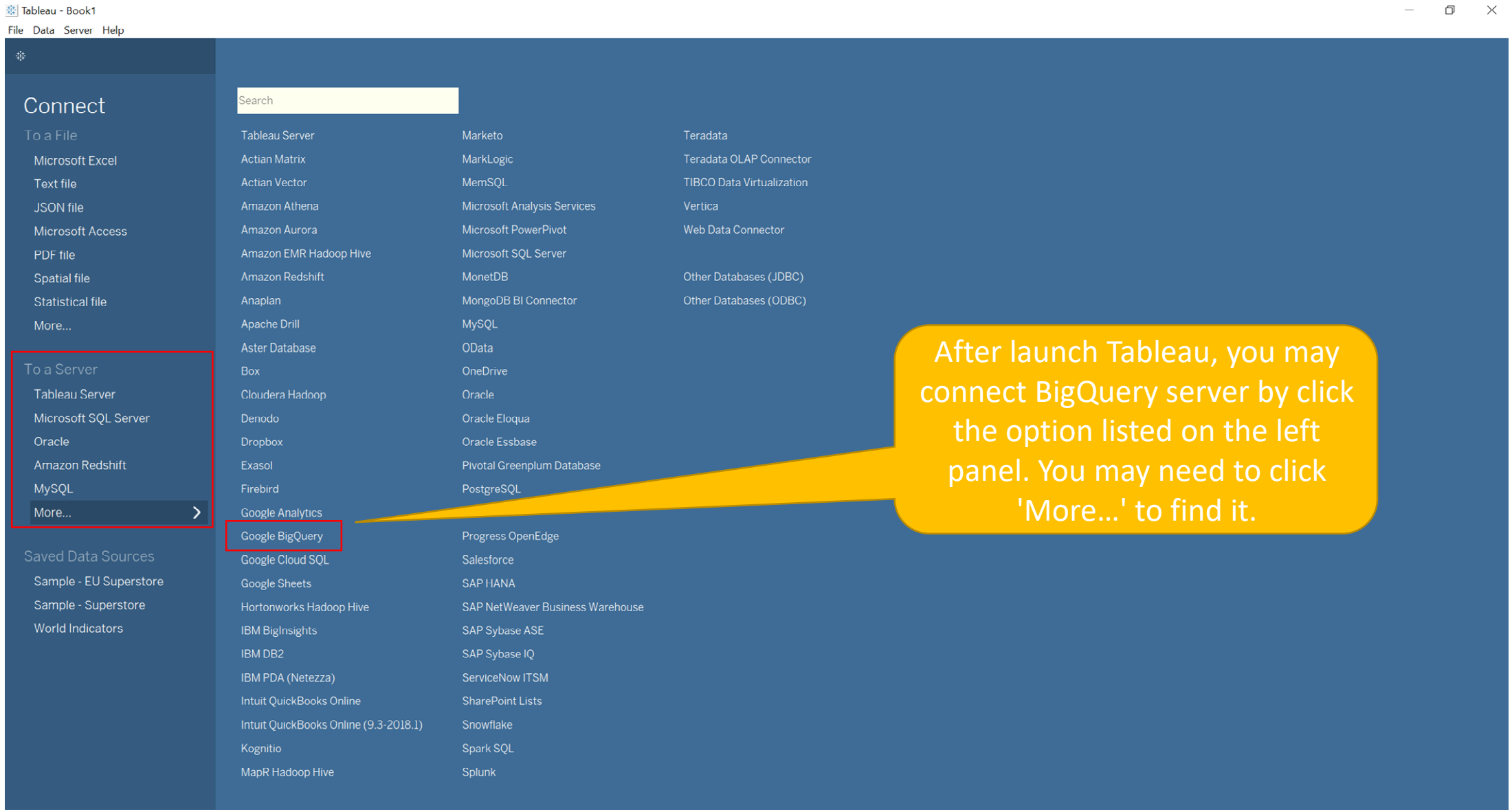
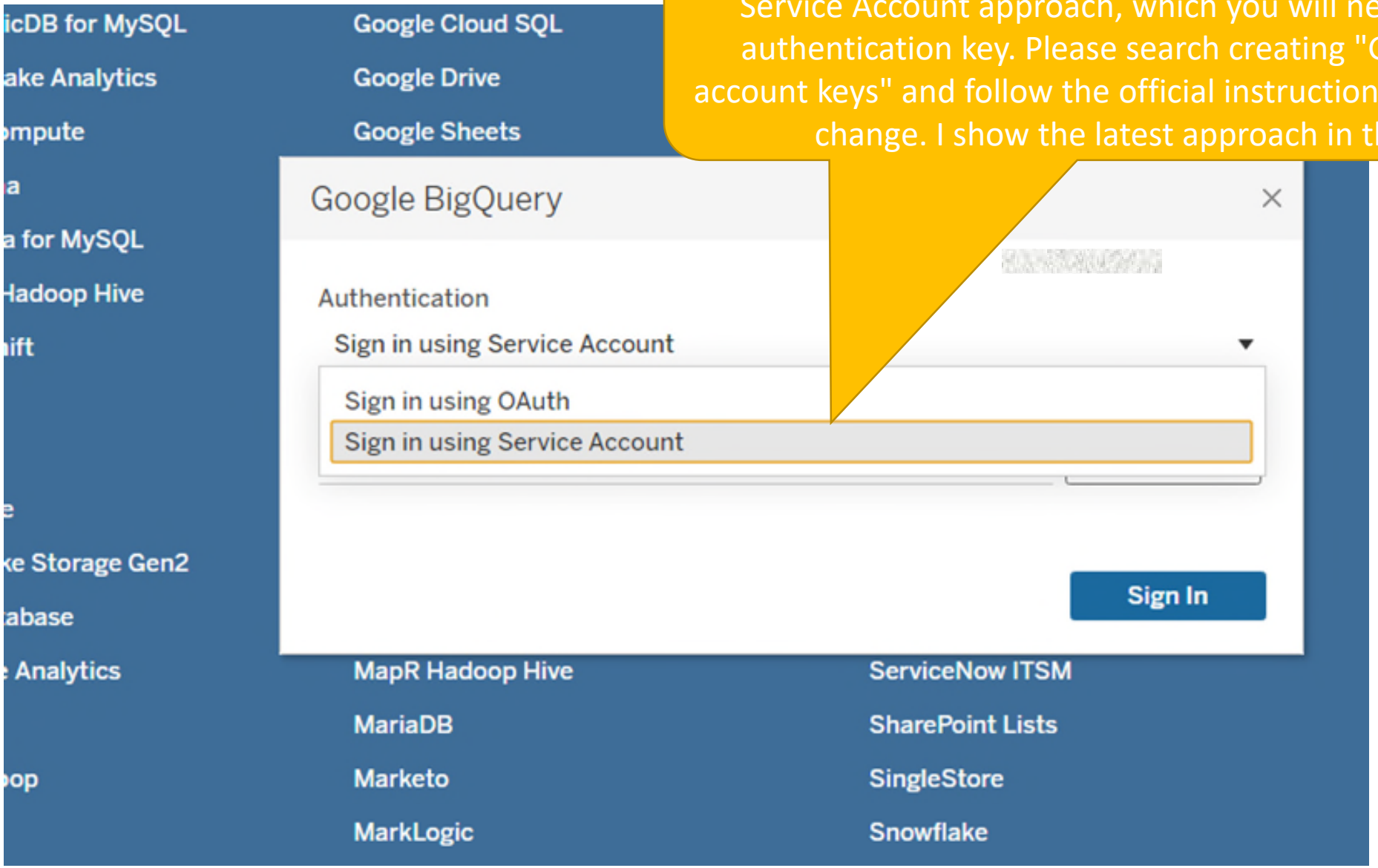


Tableau - Book1
File Data Server Help
Connect
To a File
Microsoft Excel
Text file
JSON file
Microsoft Access
PDF file
Spatial file
Statistical file
More...
To a Server
Tableau Server
Microsoft SQL Server
Oracle
Amazon Redshift
MySQL
More...
Saved Data Sources
Sample - EU Superstore
Sample - Superstore
World Indicators
Search
Tableau Server
Actian Matrix
Actian Vector
Amazon Athena
Amazon Aurora
Amazon EMR Hadoop Hive
Amazon Redshift
Anaplan
Apache Drill
Aster Database
Box
Cloudera Hadoop
Denodo
Dropbox
Exasol
Firebird
Google Analytics
Google BigQuery
Google Cloud SQL
Google Sheets
Hortonworks Hadoop Hive
IBM BigInsights
IBM DB2
IBM PDA (Netezza)
Intuit QuickBooks Online
Intuit QuickBooks Online (9.3-2018.1)
Kognitio
MapR Hadoop Hive
Marketo
MarkLogic
MemSQL
Microsoft Analysis Services
Microsoft PowerPivot
Microsoft SQL Server
MonetDB
MongoDB BI Connector
MySQL
OData
OneDrive
Oracle
Oracle Eloqua
Oracle Essbase
Pivotal Greenplum Database
PostgreSQL
Progress OpenEdge
Salesforce
SAP HANA
SAP NetWeaver Business Warehouse
SAP Sybase ASE
SAP Sybase IQ
ServiceNow ITSM
SharePoint Lists
Snowflake
Spark SQL
Splunk
Teradata
Teradata OLAP Connector
TIBCO Data Virtualization
Vertica
Web Data Connector
Other Databases (JDBC)
Other Databases (ODBC)
After launch Tableau, you may connect BigQuery server by click the option listed on the left panel. You may need to click 'More...' to find it.

There are two ways to sign in BigQuery service and I will use the Service Account approach, which you will need to create a authentication key. Please search creating "Google service account keys" and follow the official instruction as the steps may change. I show the latest approach in the video.



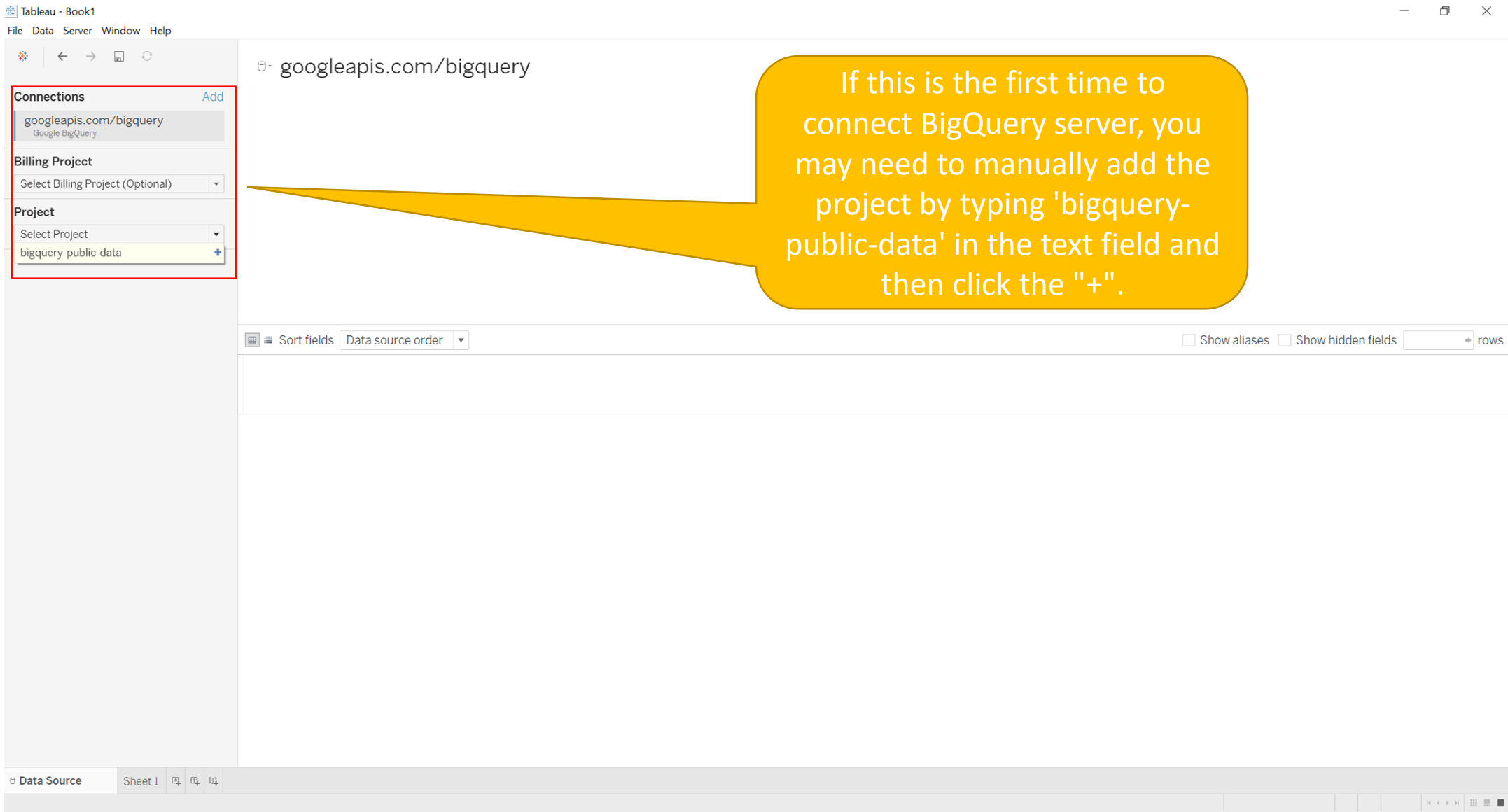
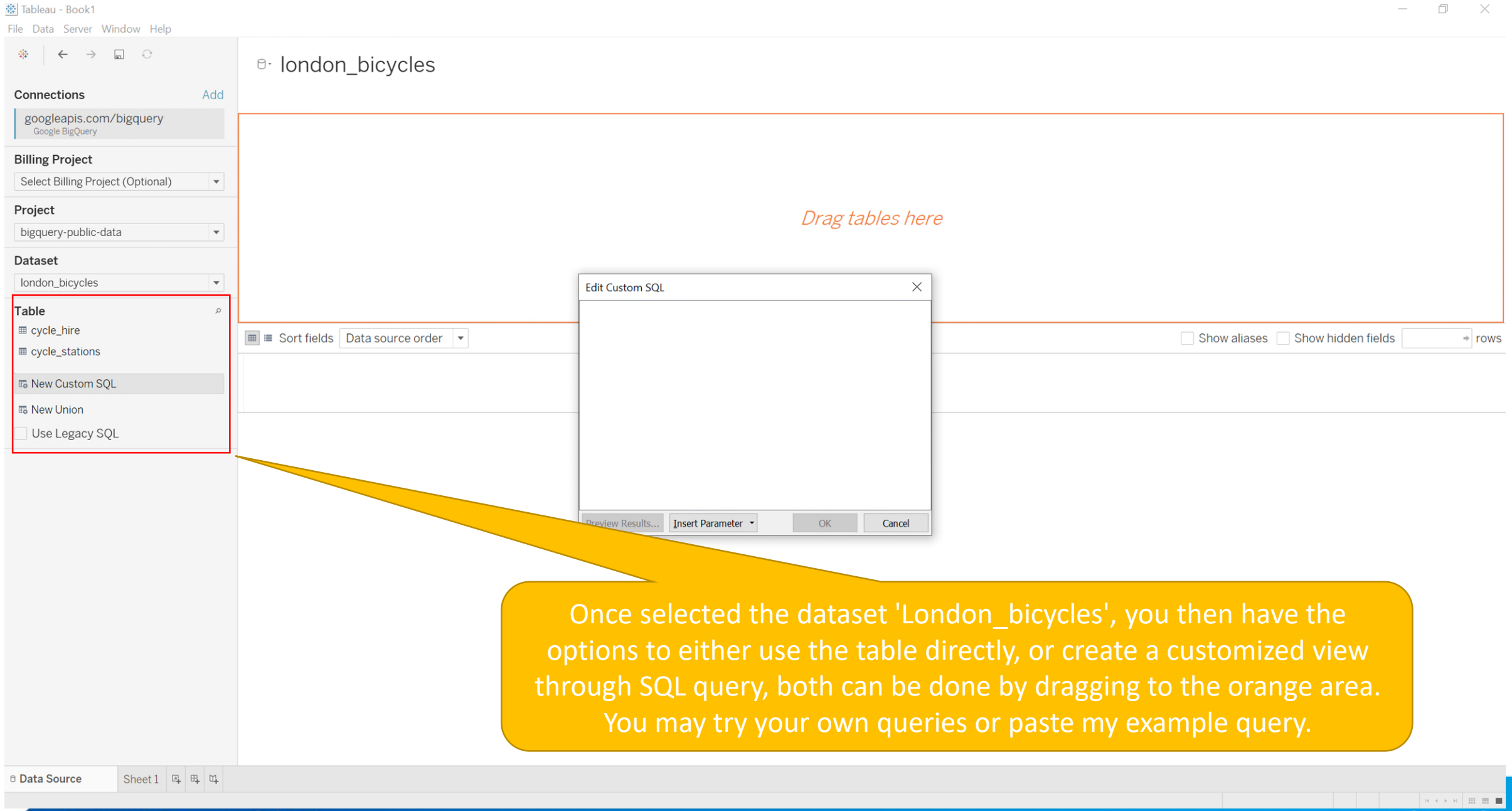


Tableau - Book1
File Data Server Window Help
Connections Add
googleapis.com/bigquery
Google BigQuery
Billing Project
Select Billing Project (Optional)
Project
Select Project
bigquery-public-data
googleapis.com/bigquery
If this is the first time to connect BigQuery server, you may need to manually add the project by typing 'bigquery-public-data' in the text field and then click the "+".
Sort fields
Data source order
Show aliases
Show hidden fields
rows
Data Source
Sheet 1

COVID-19 NYT example

```
SELECT t.state_name,AVG(t.deaths) AS deaths, AVG(m.never+m.rarely) AS avg_percent
FROM `bigquery-public-data.covid19_nyt.mask_use_by_county` AS m
JOIN `bigquery-public-data.covid19_nyt.us_counties` as c
ON m.county_fips_code = c.county_fips_code
JOIN (SELECT state_name, deaths, confirmed_cases
FROM `bigquery-public-data.covid19_nyt.us_states`
WHERE date = '2022-01-20'
ORDER BY deaths DESC
) AS t
ON t.state_name = c.state_name
GROUP BY t.state_name
ORDER BY deaths DESC
```



Once selected the dataset 'London_bicycles', you then have the options to either use the table directly, or create a customized view through SQL query, both can be done by dragging to the orange area. You may try your own queries or paste my example query.

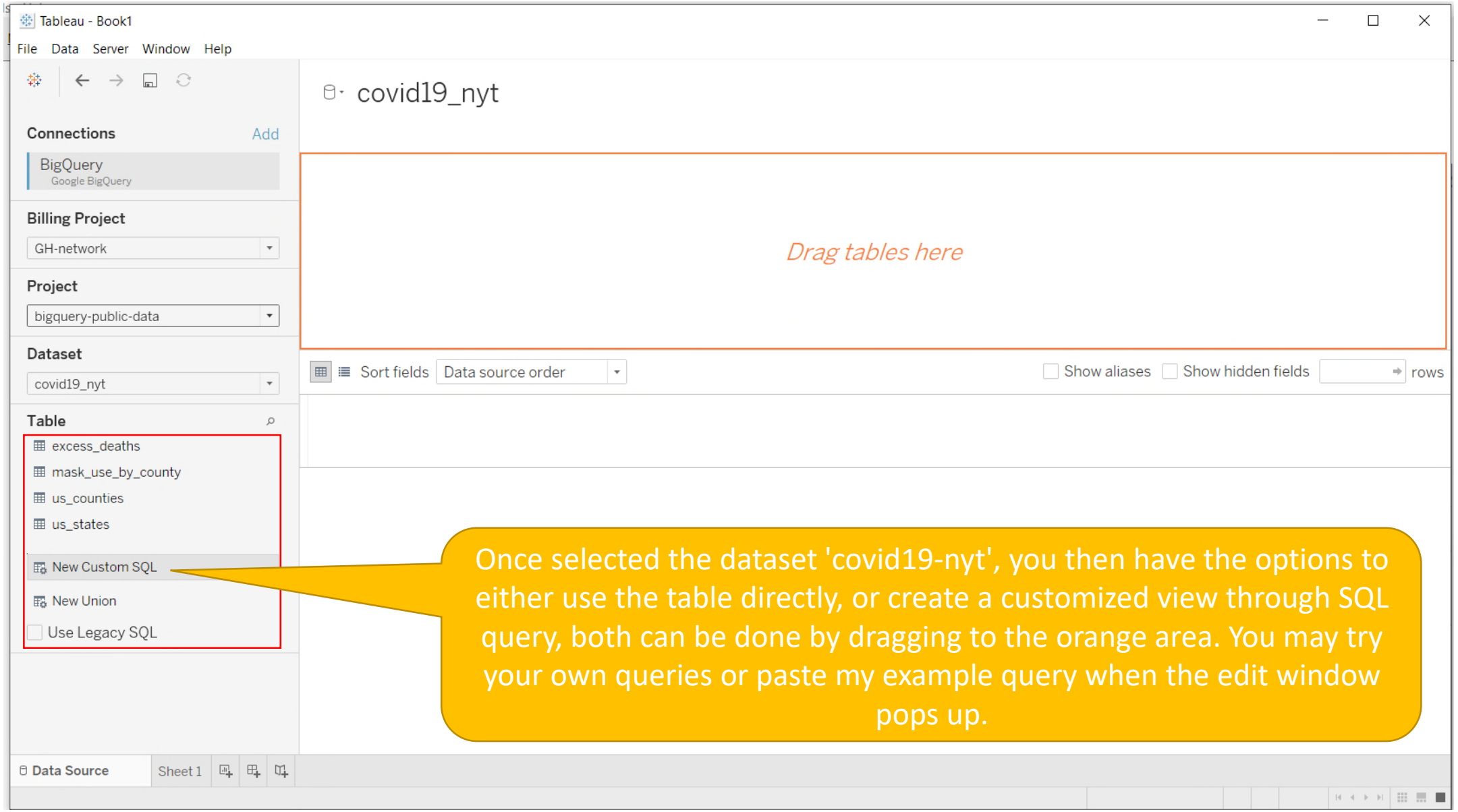


Tableau - Book1
File Data Server Window Help
covid19_nyt
Connections Add
BigQuery
Google BigQuery
Billing Project
GH-network
Project
bigquery-public-data
Dataset
covid19_nyt
Table
excess_deaths
mask_use_by_county
us_counties
us_states
New Custom SQL
New Union
Use Legacy SQL
Drag tables here
Sort fields Data source order
Show aliases Show hidden fields
rows
Once selected the dataset 'covid19-nyt', you then have the options to either use the table directly, or create a customized view through SQL query, both can be done by dragging to the orange area. You may try your own queries or paste my example query when the edit window pops up.
Data Source
Sheet 1

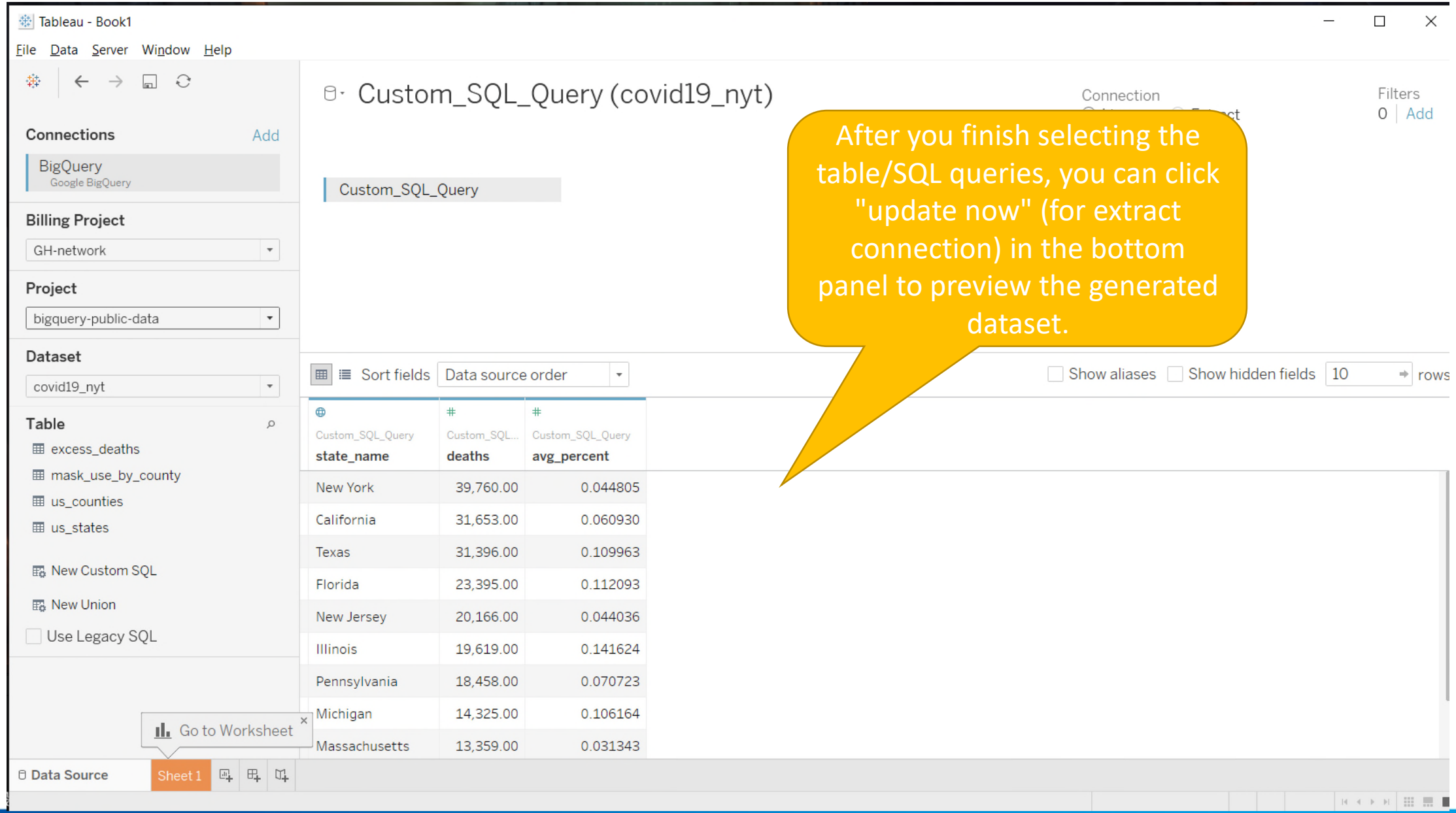
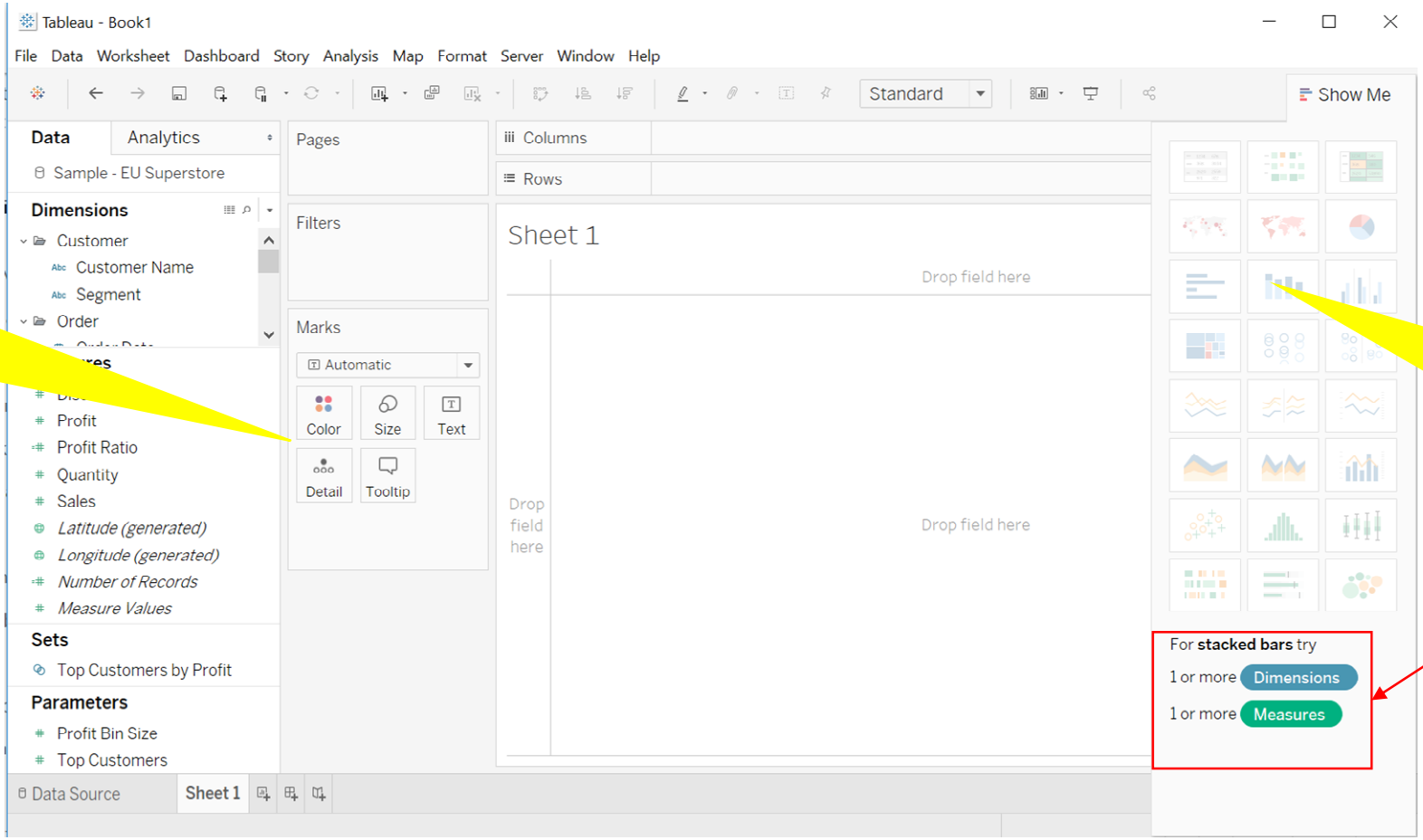


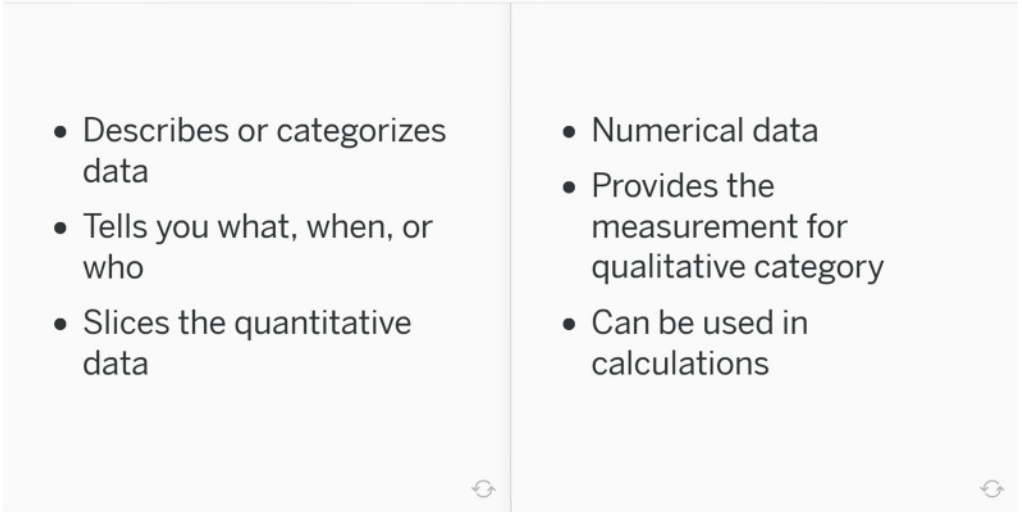
Tableau - Book1
File Data Server Window Help
Connections Add
BigQuery
Google BigQuery
Billing Project
GH-network
Project
bigquery-public-data
Dataset
covid19_nyt
Table
excess_deaths
mask_use_by_county
us_counties
us_states
New Custom SQL
New Union
Use Legacy SQL
Custom_SQL_Query (covid19_nyt)
Custom_SQL_Query
Connection
Filters
0 Add
After you finish selecting the table/SQL queries, you can click "update now" (for extract connection) in the bottom panel to preview the generated dataset.
Sort fields Data source order
Show aliases Show hidden fields 10 rows
Custom_SQL_Query state_name
Custom_SQL... deaths
Custom_SQL_Query avg_percent
New York 39,760.00 0.044805
California 31,653.00 0.060930
Texas 31,396.00 0.109963
Florida 23,395.00 0.112093
New Jersey 20,166.00 0.044036
Illinois 19,619.00 0.141624
Pennsylvania 18,458.00 0.070723
Michigan 14,325.00 0.106164
Massachusetts 13,359.00 0.031343
Go to Worksheet
Data Source
Sheet 1

Drag dimensions or measures to make changes, such as color, size, to the selected chart.



Move your mouse to the desired chart type. It shows required measures and dimensions.

- When users connect to Tableau, the data fields in their data set are automatically assigned a role and a type.
- In Tableau, quantitative fields are referred to as Measures, and qualitative fields are referred to as Dimensions.



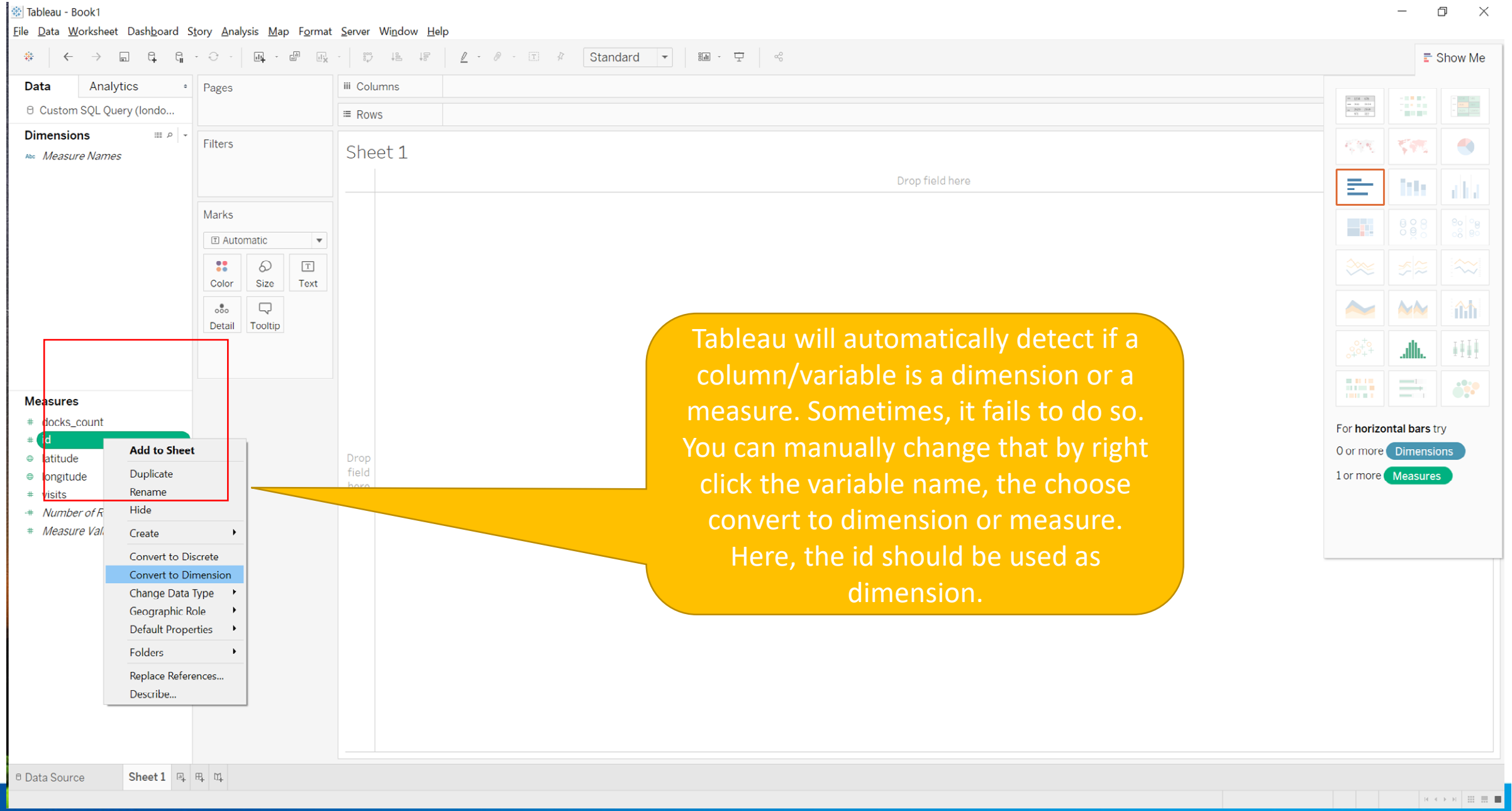


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Custom SQL Query (londo...
Dimensions
Measure Names
Pages
Filters
Marks
Automatic
Color
Size
Text
Detail
Tooltip
Columns
Rows
Sheet 1
Drop field here
Measures
docks_count
id
latitude
longitude
visits
Number of R
Measure Val
Add to Sheet
Duplicate
Rename
Hide
Create
Convert to Discrete
Convert to Dimension
Change Data Type
Geographic Role
Default Properties
Folders
Replace References...
Describe...
For horizontal bars try
0 or more Dimensions
1 or more Measures
Data Source
Sheet 1
Tableau will automatically detect if a column/variable is a dimension or a measure. Sometimes, it fails to do so. You can manually change that by right click the variable name, the choose convert to dimension or measure. Here, the id should be used as dimension.

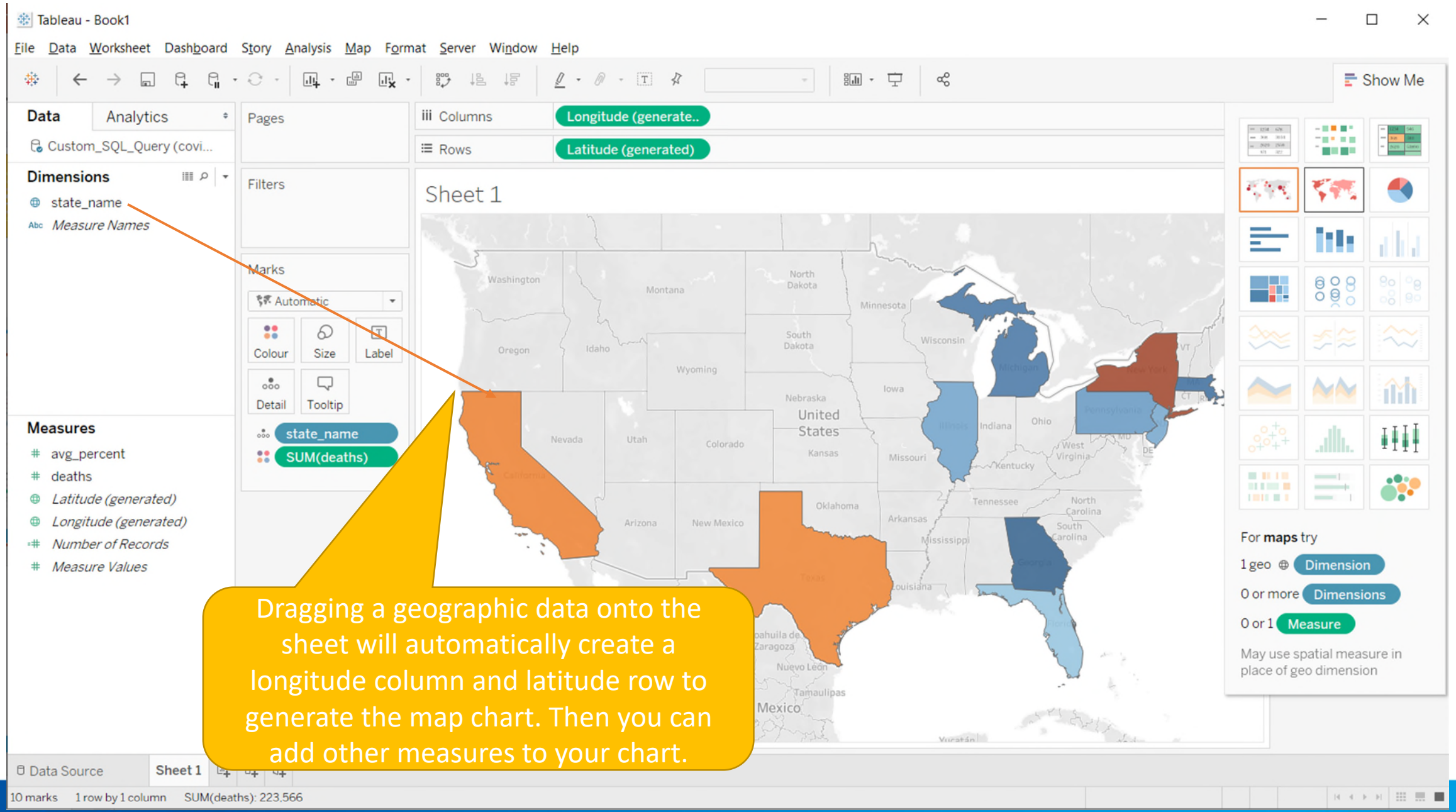
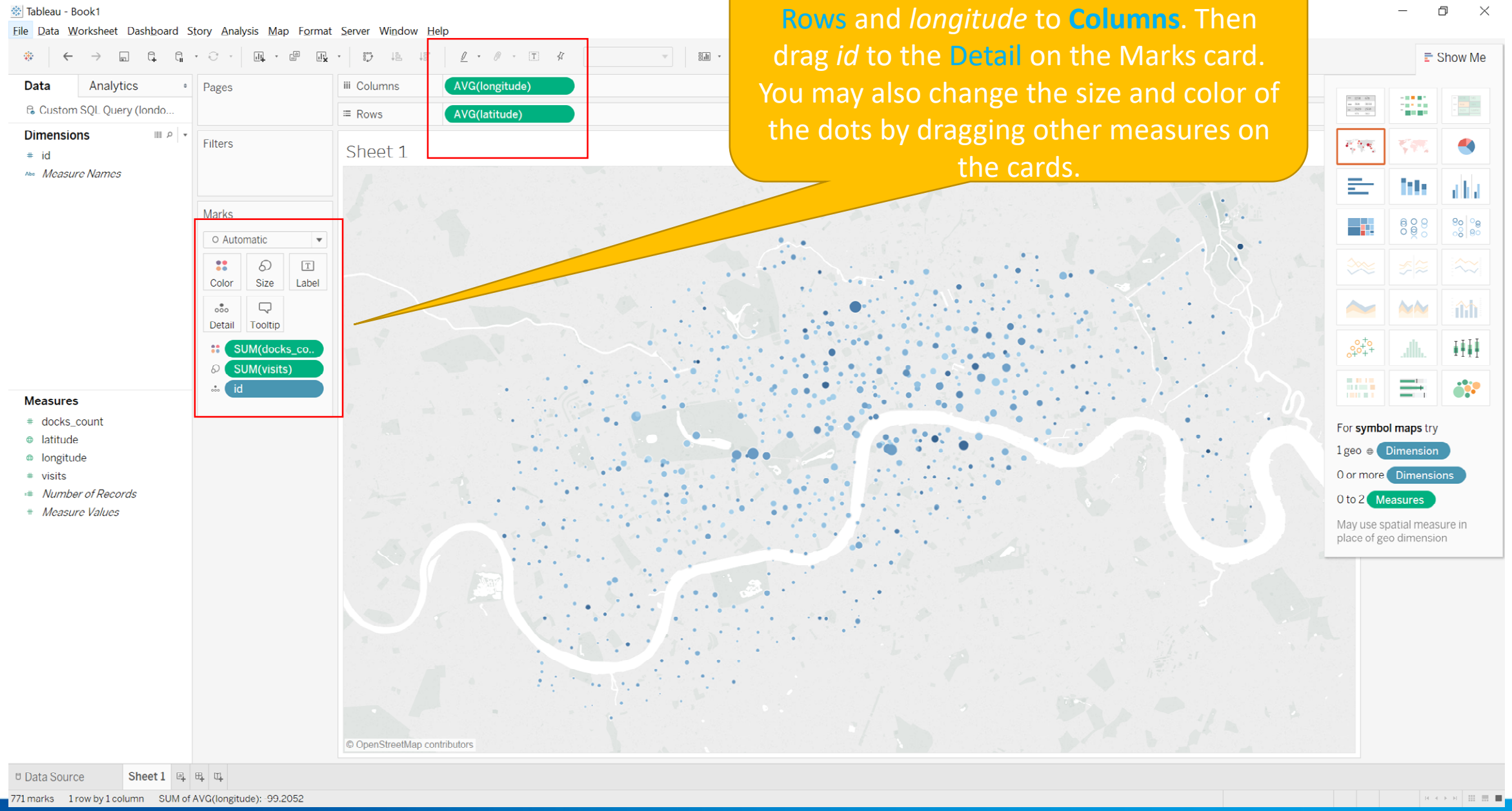


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Custom_SQL_Query (covi...
Dimensions
state_name
Measure Names
Measures
avg_percent
deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Colour
Size
Label
Detail
Tooltip
state_name
SUM(deaths)
Columns
Longitude (generate..
Rows
Latitude (generated)
Sheet 1
Washington
Montana
North Dakota
Minnesota
Oregon
Idaho
South Dakota
Wisconsin
Wyoming
Nebraska
Iowa
United States
Nevada
Utah
Colorado
Kansas
Missouri
Illinois
Indiana
Ohio
California
Arizona
New Mexico
Oklahoma
Arkansas
Kentucky
West Virginia
Tennessee
North Carolina
South Carolina
Mississippi
Georgia
Louisiana
Texas
Nuevo León
Tamaulipas
Mexico
For maps try
1 geo Dimension
0 or more Dimensions
0 or 1 Measure
May use spatial measure in place of geo dimension
Dragging a geographic data onto the sheet will automatically create a longitude column and latitude row to generate the map chart. Then you can add other measures to your chart.
Data Source
Sheet 1
10 marks
1 row by 1 column
SUM(deaths): 223,566


To create a map chart, drag *latitude* to Rows and *longitude* to **Columns**. Then drag *id* to the Detail on the Marks card. You may also change the size and color of the dots by dragging other measures on the cards.
Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Data
Analytics
Custom SQL Query (londo...
Dimensions
id
Measure Names
Measures
docks_count
latitude
longitude
visits
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
SUM(docks_co..
SUM(visits)
id
Columns
AVG(longitude)
Rows
AVG(latitude)
Sheet 1
Show Me
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
© OpenStreetMap contributors
Data Source
Sheet 1
771 marks
1 row by 1 column
SUM of AVG(longitude): 99.2052

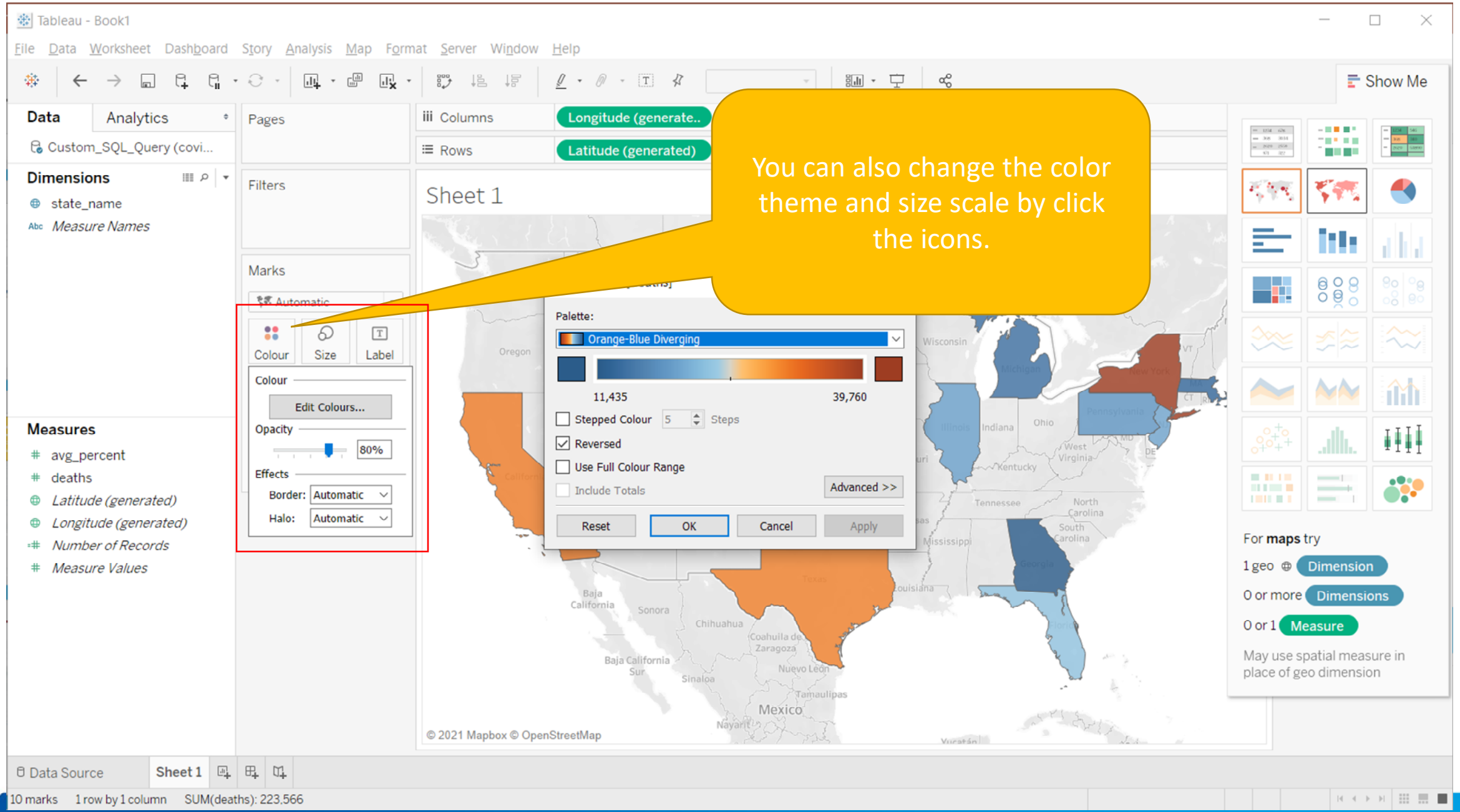
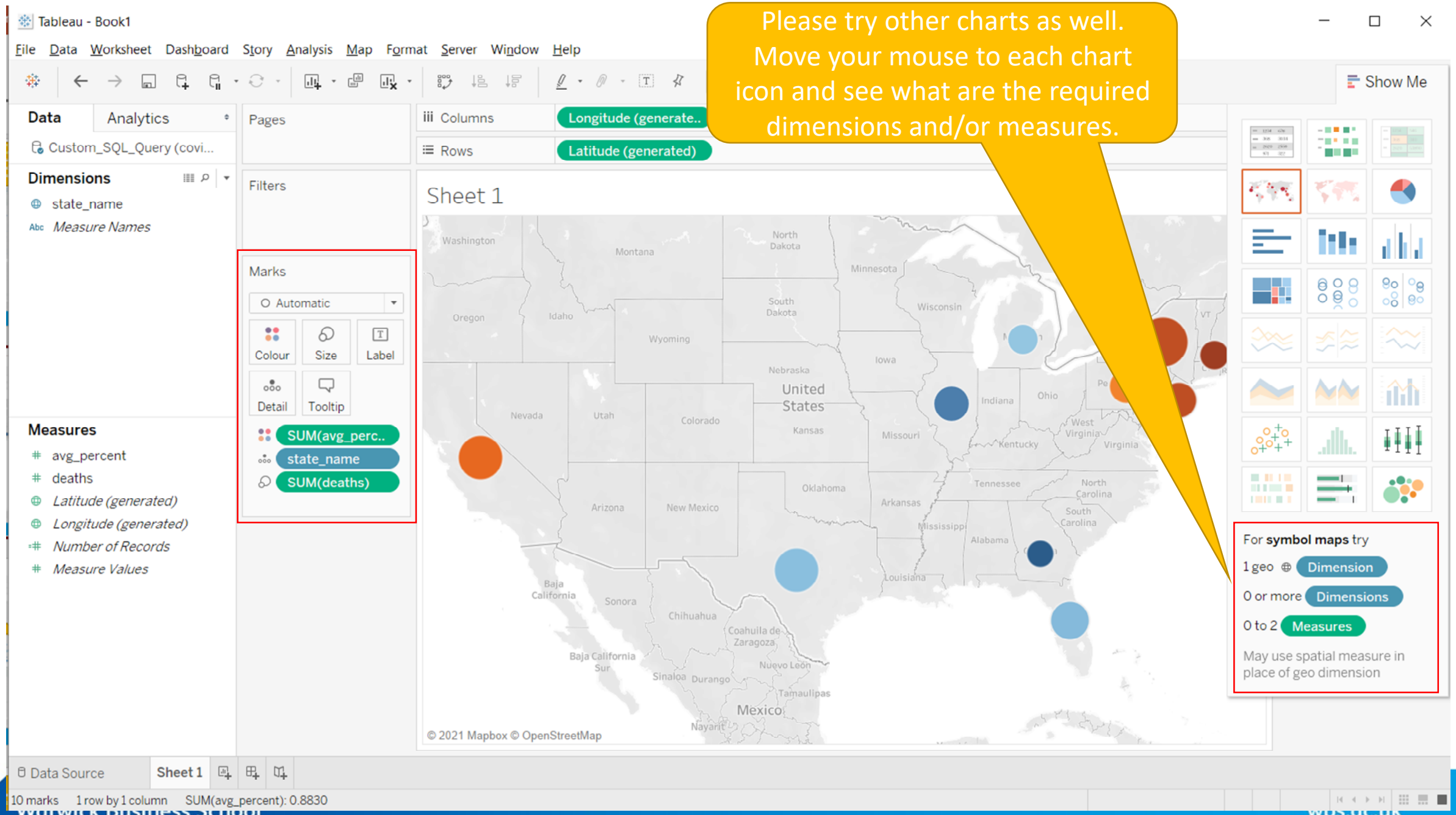
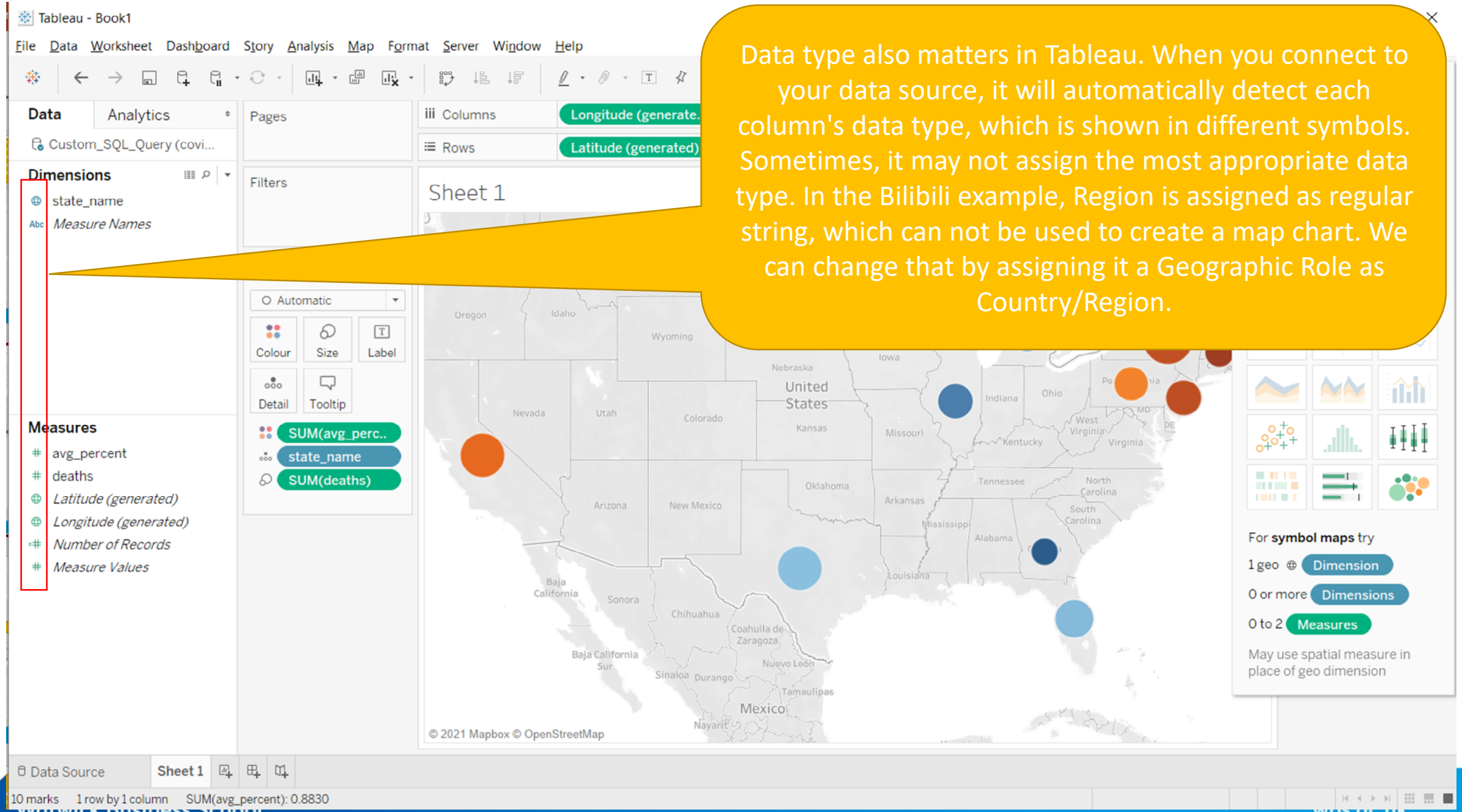


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Custom_SQL_Query (covi...
Dimensions
state_name
Measure Names
Measures
avg_percent
deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Colour
Size
Label
Colour
Edit Colours...
Opacity
80%
Effects
Border: Automatic
Halo: Automatic
Columns
Longitude (generate..
Rows
Latitude (generated)
Sheet 1
You can also change the color theme and size scale by click the icons.
Palette:
Orange-Blue Diverging
11,435
39,760
Stepped Colour
5
Steps
Reversed
Use Full Colour Range
Include Totals
Advanced >>
Reset
OK
Cancel
Apply
For maps try
1 geo Dimension
0 or more Dimensions
0 or 1 Measure
May use spatial measure in place of geo dimension
© 2021 Mapbox © OpenStreetMap
Data Source
Sheet 1
10 marks
1 row by 1 column
SUM(deaths): 223,566


Please try other charts as well. Move your mouse to each chart icon and see what are the required dimensions and/or measures.
Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Custom_SQL_Query (covi...
Dimensions
state_name
Measure Names
Measures
avg_percent
deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Colour
Size
Label
Detail
Tooltip
SUM(avg_perc..
state_name
SUM(deaths)
Columns
Longitude (generate..
Rows
Latitude (generated)
Sheet 1
Washington
Montana
North Dakota
Minnesota
Oregon
Idaho
South Dakota
Wisconsin
Wyoming
Nebraska
Iowa
United States
Nevada
Utah
Colorado
Kansas
Missouri
Indiana
Ohio
West Virginia
Kentucky
Virginia
Oklahoma
Arkansas
Tennessee
North Carolina
South Carolina
Arizona
New Mexico
Mississippi
Alabama
Louisiana
Baja California
Sonora
Chihuahua
Coahuila de Zaragoza
Baja California Sur
Nuevo León
Sinaloa
Durango
Tamaulipas
Mexico
Nayarit
© 2021 Mapbox © OpenStreetMap
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
Data Source
Sheet 1
10 marks
1 row by 1 column
SUM(avg_percent): 0.8830



Data type also matters in Tableau. When you connect to your data source, it will automatically detect each column's data type, which is shown in different symbols. Sometimes, it may not assign the most appropriate data type. In the Bilibili example, Region is assigned as regular string, which can not be used to create a map chart. We can change that by assigning it a Geographic Role as Country/Region.

Working with multiple data sources

- You may add more than one dataset.
  - Different sources
  - Different formats
- Combine different datasets via Join.
  - Automatically detect the relationships
  - Manually edit the relationships

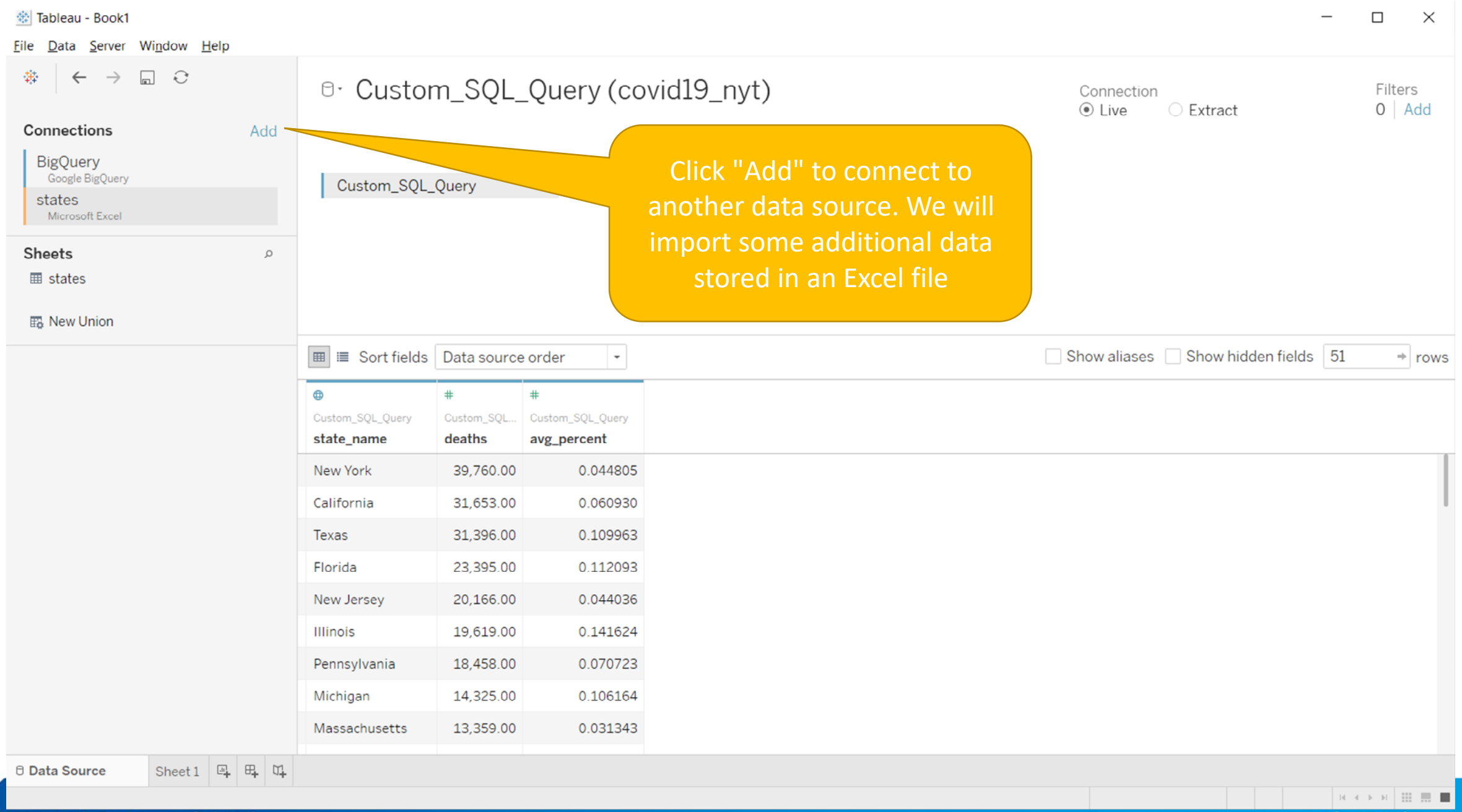


Tableau - Book1
File Data Server Window Help
Custom_SQL_Query (covid19_nyt)
Connection
Live
Extract
Filters
0 Add
Connections
Add
BigQuery
Google BigQuery
states
Microsoft Excel
Custom_SQL_Query
Click "Add" to connect to another data source. We will import some additional data stored in an Excel file
Sheets
states
New Union
Sort fields
Data source order
Show aliases
Show hidden fields
51
rows
Custom_SQL_Query
state_name
Custom_SQL...
deaths
Custom_SQL_Query
avg_percent
New York 39,760.00 0.044805
California 31,653.00 0.060930
Texas 31,396.00 0.109963
Florida 23,395.00 0.112093
New Jersey 20,166.00 0.044036
Illinois 19,619.00 0.141624
Pennsylvania 18,458.00 0.070723
Michigan 14,325.00 0.106164
Massachusetts 13,359.00 0.031343
Data Source
Sheet 1

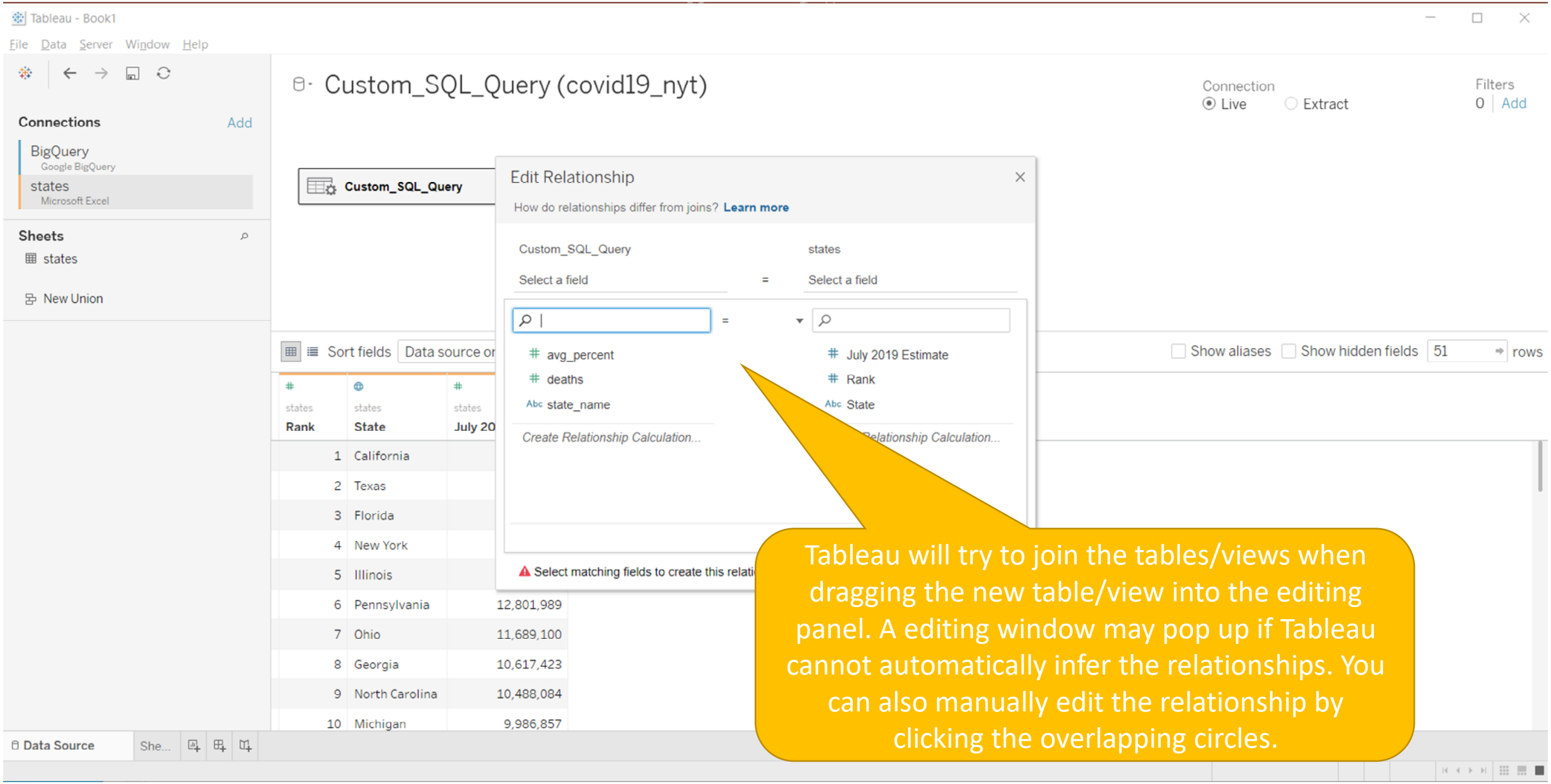


Tableau - Book1
File Data Server Window Help
Custom_SQL_Query (covid19_nyt)
Connection
Live
Extract
Filters
0 Add
Connections
Add
BigQuery
Google BigQuery
states
Microsoft Excel
Sheets
states
New Union
Custom_SQL_Query
Edit Relationship
How do relationships differ from joins? Learn more
Custom_SQL_Query
states
Select a field
=
Select a field
avg_percent
deaths
state_name
July 2019 Estimate
Rank
State
Create Relationship Calculation...
Select matching fields to create this relati
Sort fields
Data source or
Show aliases
Show hidden fields
51
rows
states Rank
states State
states July 20
1 California
2 Texas
3 Florida
4 New York
5 Illinois
6 Pennsylvania 12,801,989
7 Ohio 11,689,100
8 Georgia 10,617,423
9 North Carolina 10,488,084
10 Michigan 9,986,857
Data Source
She...
Tableau will try to join the tables/views when dragging the new table/view into the editing panel. A editing window may pop up if Tableau cannot automatically infer the relationships. You can also manually edit the relationship by clicking the overlapping circles.

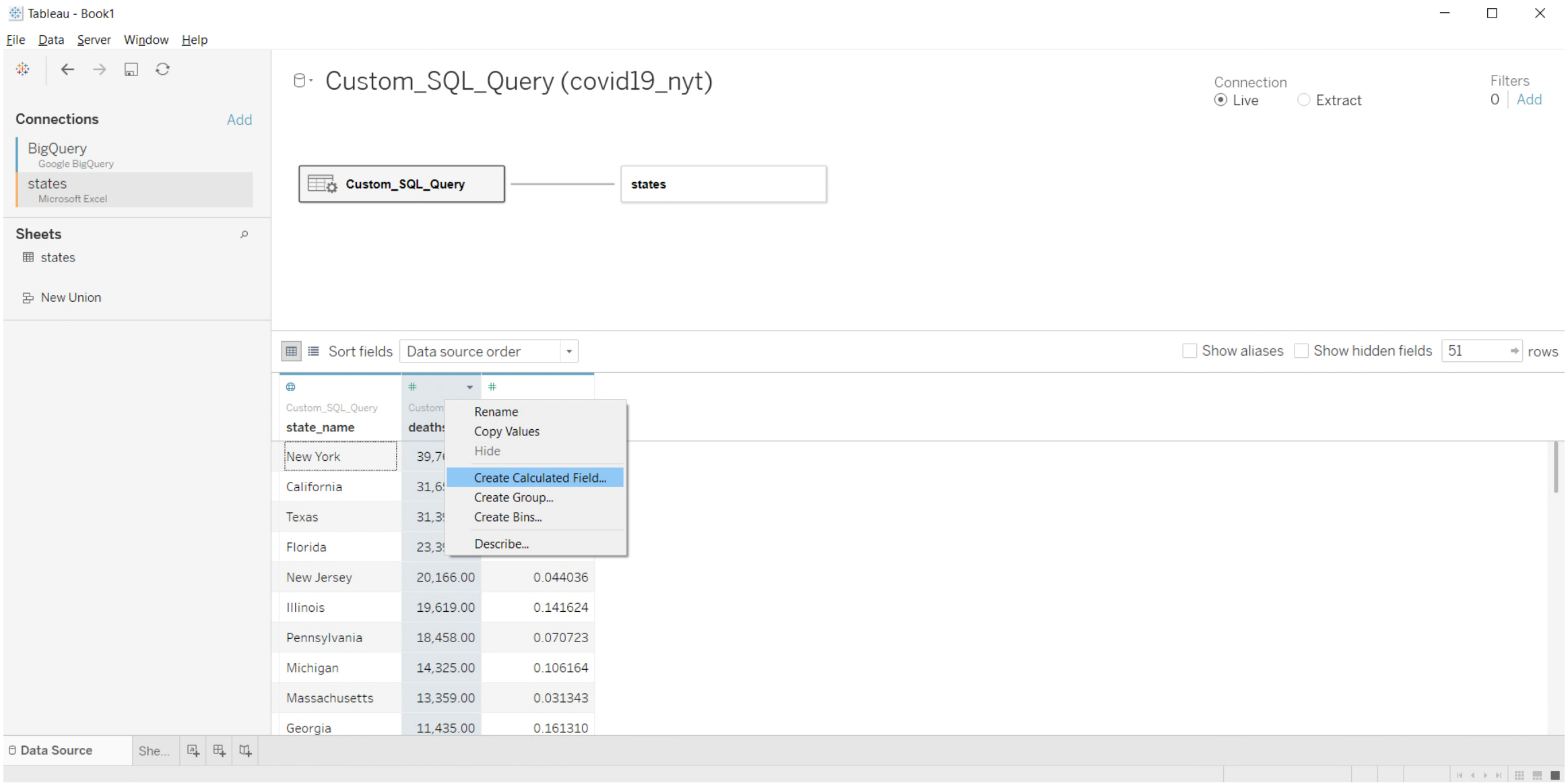


Tableau - Book1
File Data Server Window Help
Connections Add
BigQuery
Google BigQuery
states
Microsoft Excel
Sheets
states
New Union
Custom_SQL_Query (covid19_nyt)
Connection
Live
Extract
Filters
0 Add
Custom_SQL_Query
states
Sort fields
Data source order
Show aliases
Show hidden fields
51
rows
Custom_SQL_Query
state_name
Rename
Copy Values
Hide
Create Calculated Field...
Create Group...
Create Bins...
Describe...
New York
California
Texas
Florida
New Jersey 20,166.00 0.044036
Illinois 19,619.00 0.141624
Pennsylvania 18,458.00 0.070723
Michigan 14,325.00 0.106164
Massachusetts 13,359.00 0.031343
Georgia 11,435.00 0.161310
Data Source

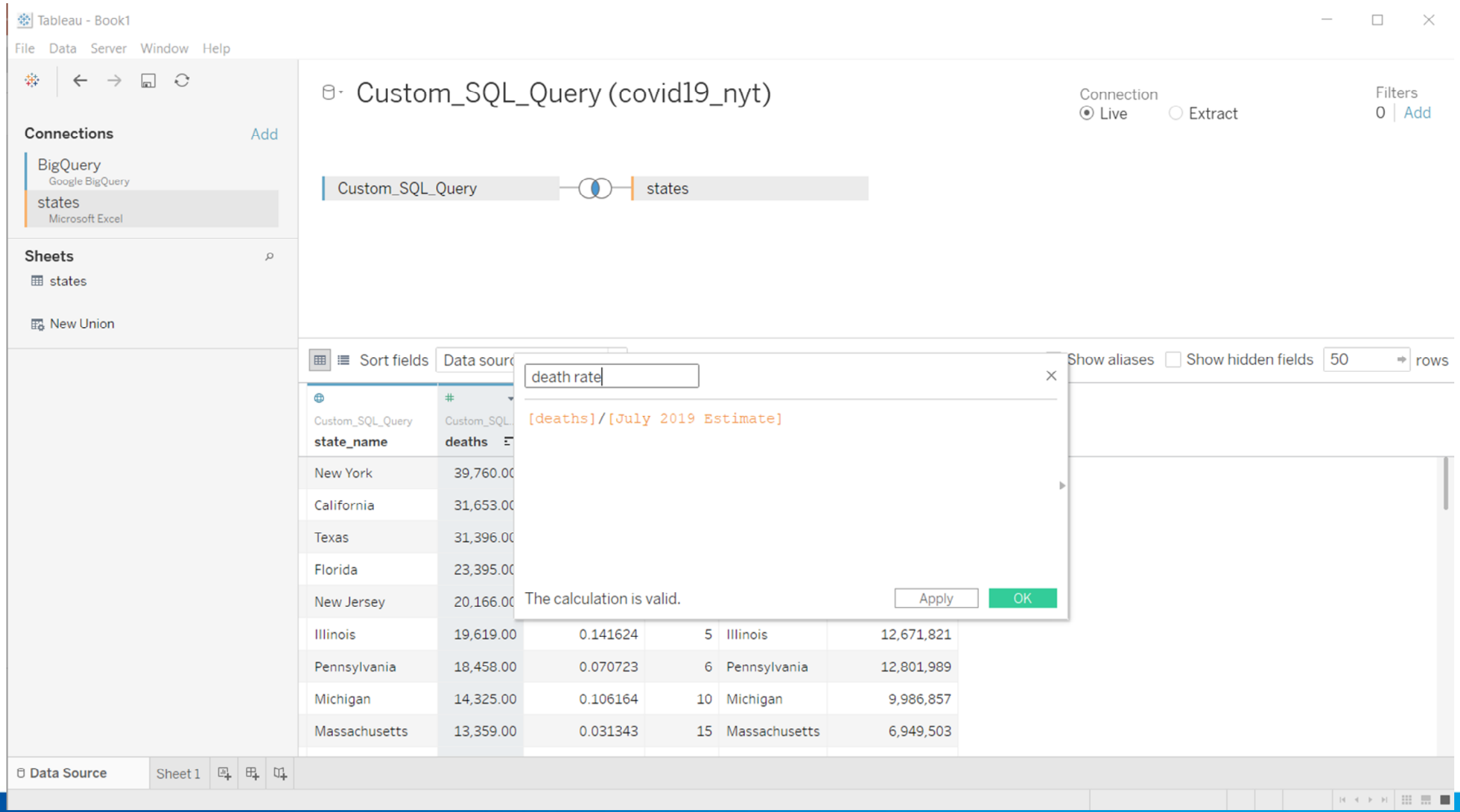


Tableau - Book1
File Data Server Window Help
Connections Add
BigQuery
Google BigQuery
states
Microsoft Excel
Sheets
states
New Union
Custom_SQL_Query (covid19_nyt)
Connection
Live
Extract
Filters
0 Add
Custom_SQL_Query
states
Sort fields
Show aliases
Show hidden fields
50
rows
death rate
[deaths]/[July 2019 Estimate]
The calculation is valid.
Apply
OK
Custom_SQL_Query
state_name
deaths
New York 39,760.00
California 31,653.00
Texas 31,396.00
Florida 23,395.00
New Jersey 20,166.00
Illinois 19,619.00 0.141624 5 Illinois 12,671,821
Pennsylvania 18,458.00 0.070723 6 Pennsylvania 12,801,989
Michigan 14,325.00 0.106164 10 Michigan 9,986,857
Massachusetts 13,359.00 0.031343 15 Massachusetts 6,949,503
Data Source
Sheet 1

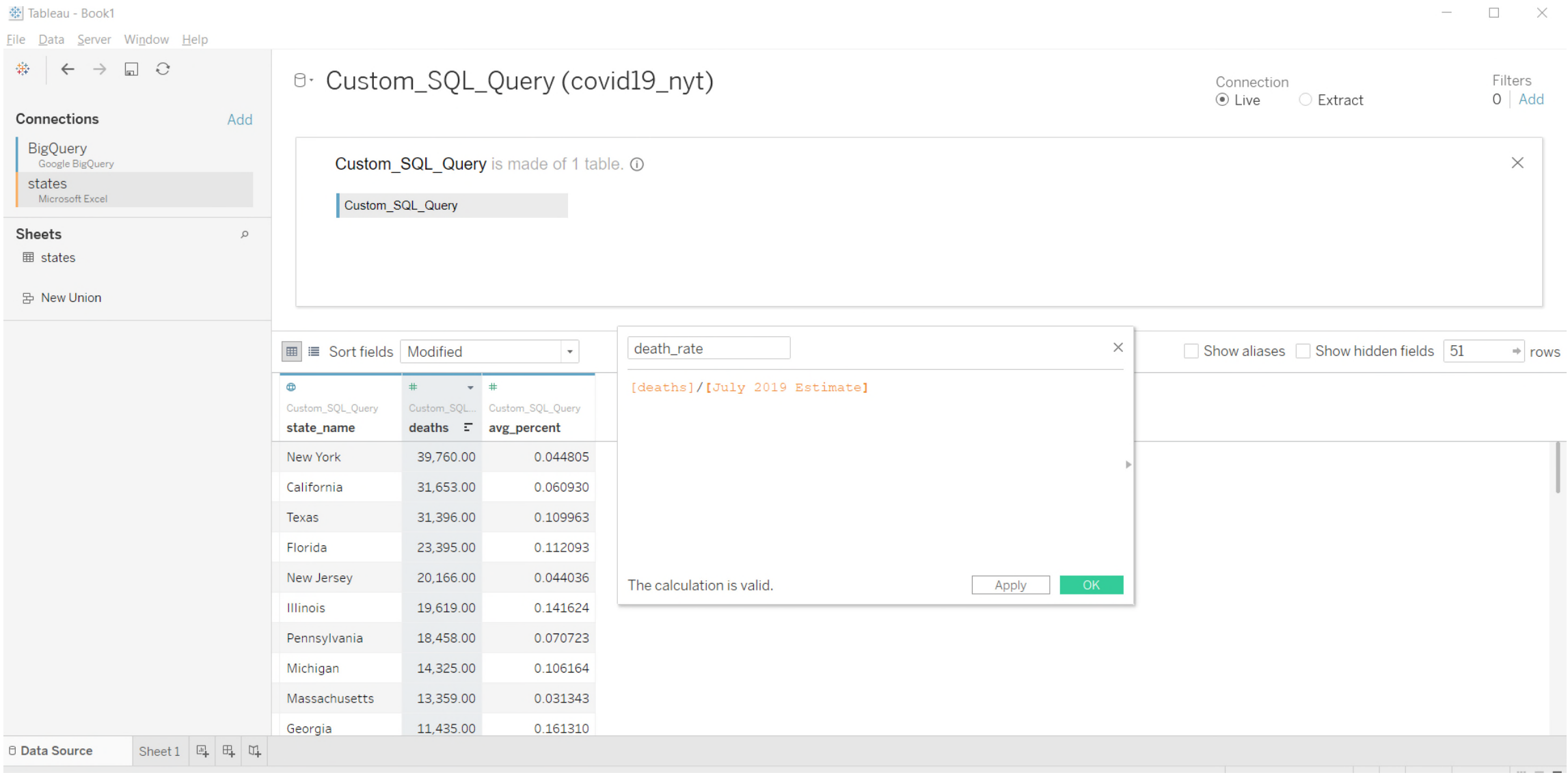


Tableau - Book1
File Data Server Window Help
Connections Add
BigQuery
Google BigQuery
states
Microsoft Excel
Sheets
states
New Union
Custom_SQL_Query (covid19_nyt)
Connection
Live
Extract
Filters
0 Add
Custom_SQL_Query is made of 1 table.
Custom_SQL_Query
Sort fields Modified
Show aliases Show hidden fields 51 rows
Custom_SQL_Query state_name
Custom_SQL... deaths
Custom_SQL_Query avg_percent
New York 39,760.00 0.044805
California 31,653.00 0.060930
Texas 31,396.00 0.109963
Florida 23,395.00 0.112093
New Jersey 20,166.00 0.044036
Illinois 19,619.00 0.141624
Pennsylvania 18,458.00 0.070723
Michigan 14,325.00 0.106164
Massachusetts 13,359.00 0.031343
Georgia 11,435.00 0.161310
death_rate
[deaths]/[July 2019 Estimate]
The calculation is valid.
Apply
OK
Data Source
Sheet 1

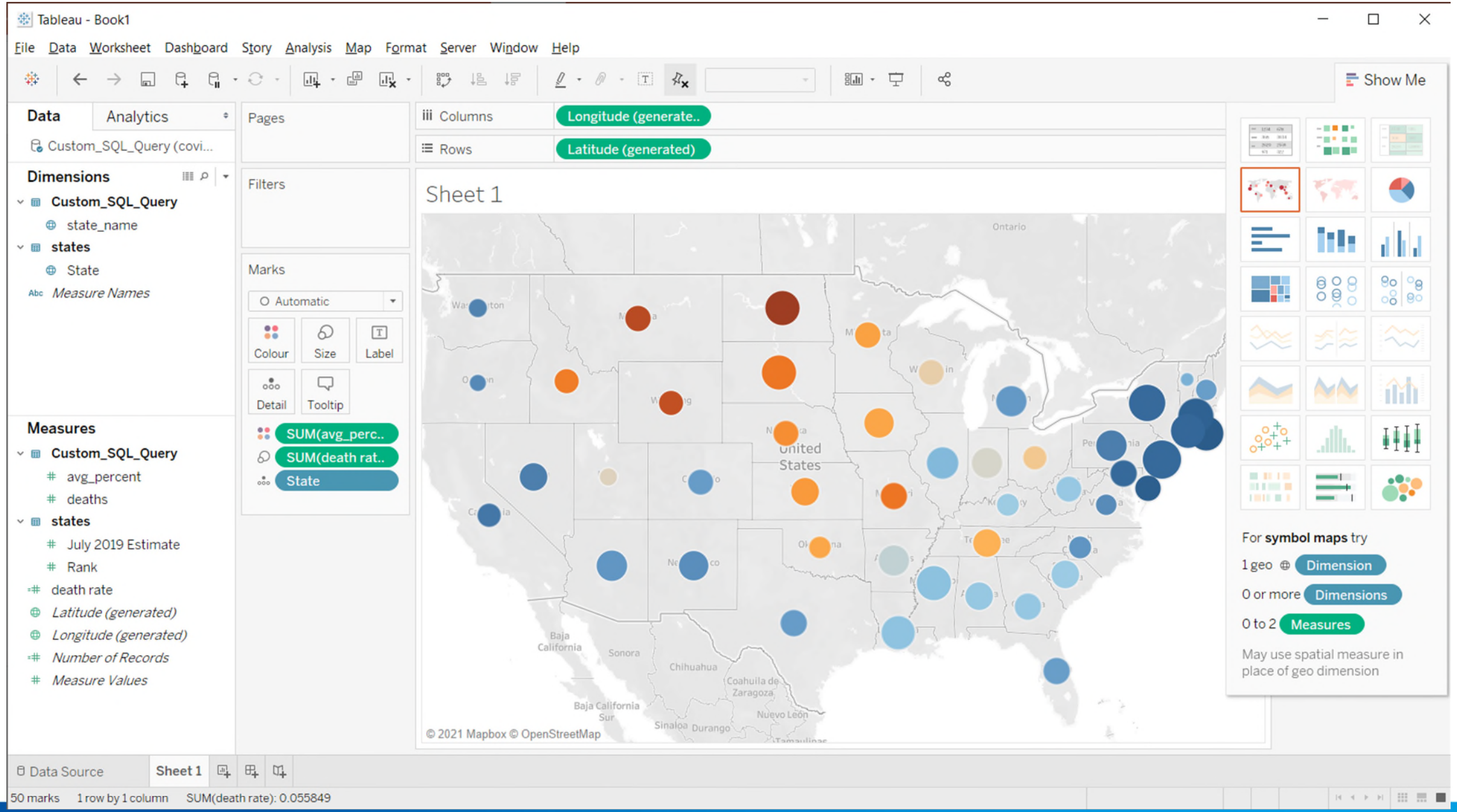
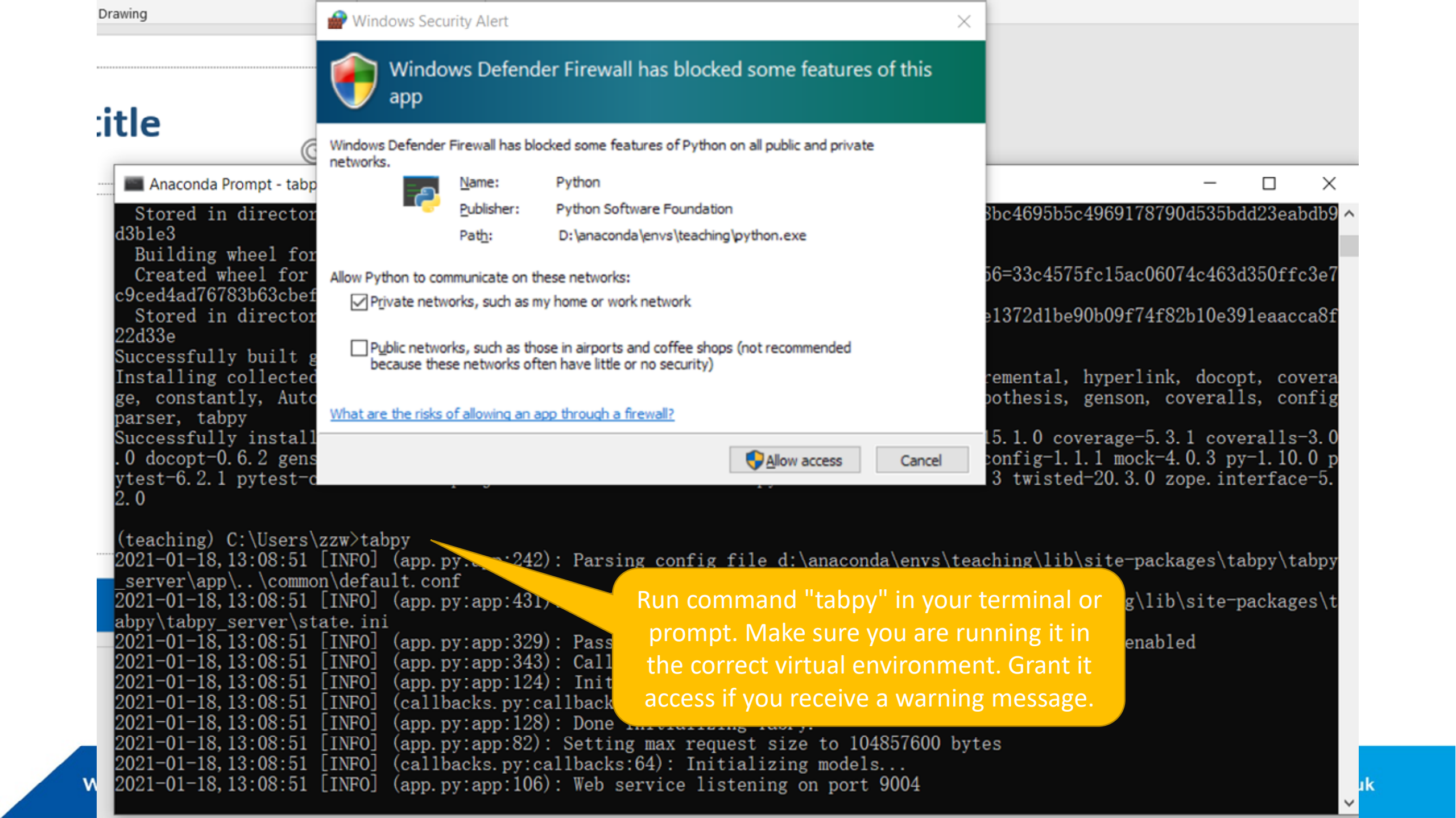


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Custom_SQL_Query (covi...
Dimensions
Custom_SQL_Query
state_name
states
State
Measure Names
Measures
Custom_SQL_Query
avg_percent
deaths
states
July 2019 Estimate
Rank
death rate
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Colour
Size
Label
Detail
Tooltip
SUM(avg_perc..
SUM(death rat..
State
Columns
Longitude (generate..
Rows
Latitude (generated)
Sheet 1
Ontario
United States
Baja California
Sonora
Chihuahua
Coahuila de Zaragoza
Baja California Sur
Nuevo León
Sinaloa
Durango
© 2021 Mapbox © OpenStreetMap
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
Data Source
Sheet 1
50 marks
1 row by 1 column
SUM(death rate): 0.055849

Connecting Tableau with Python Script

- Tableau also supports running script written in Python.
  - Connect Python backend.
  - Embed Python script as calculation.
  - pip install tabpy


Drawing
Windows Security Alert
Windows Defender Firewall has blocked some features of this app
Windows Defender Firewall has blocked some features of Python on all public and private networks.
Name: Python
Publisher: Python Software Foundation
Path: D:\anaconda\envs\teaching\python.exe
Allow Python to communicate on these networks:
Private networks, such as my home or work network
Public networks, such as those in airports and coffee shops (not recommended because these networks often have little or no security)
What are the risks of allowing an app through a firewall?
Allow access
Cancel
Anaconda Prompt - tabp
(teaching) C:\Users\zzw>tabpy
2021-01-18,13:08:51 [INFO] (app.py:app:242): Parsing config file d:\anaconda\envs\teaching\lib\site-packages\tabpy\tabpy_server\app\..\common\default.conf
2021-01-18,13:08:51 [INFO] (app.py:app:431)
abpy\tabpy_server\state.ini
2021-01-18,13:08:51 [INFO] (app.py:app:329): Pass
2021-01-18,13:08:51 [INFO] (app.py:app:343): Call
2021-01-18,13:08:51 [INFO] (app.py:app:124): Init
2021-01-18,13:08:51 [INFO] (callbacks.py:callback
2021-01-18,13:08:51 [INFO] (app.py:app:128): Done
2021-01-18,13:08:51 [INFO] (app.py:app:82): Setting max request size to 104857600 bytes
2021-01-18,13:08:51 [INFO] (callbacks.py:callbacks:64): Initializing models...
2021-01-18,13:08:51 [INFO] (app.py:app:106): Web service listening on port 9004
Run command "tabpy" in your terminal or prompt. Make sure you are running it in the correct virtual environment. Grant it access if you receive a warning message.

localhost:9004

TabPy Server Info:


```
{
    "description": "",
    "creation_time": "0",
    "state_path": "d:\\anaconda\\envs\\teaching\\lib\\site-packages\\tabpy\\tabpy_server",
    "server_version": "2.3.1",
    "name": "TabPy Server",
    "versions": {
        "v1": {
            "features": {}
        }
    }
}
```


You can check if TabPy is running by typing "localhost:9004" in your web browser. If you see similar webpage, it is running properly.

Deployed Models:


```
{}
```


Useful links:

- TabPy Documentation
- TabPy Source Code
- TabPy PyPi
- Tableau Sci-Fi - Advanced Analytics Team blog

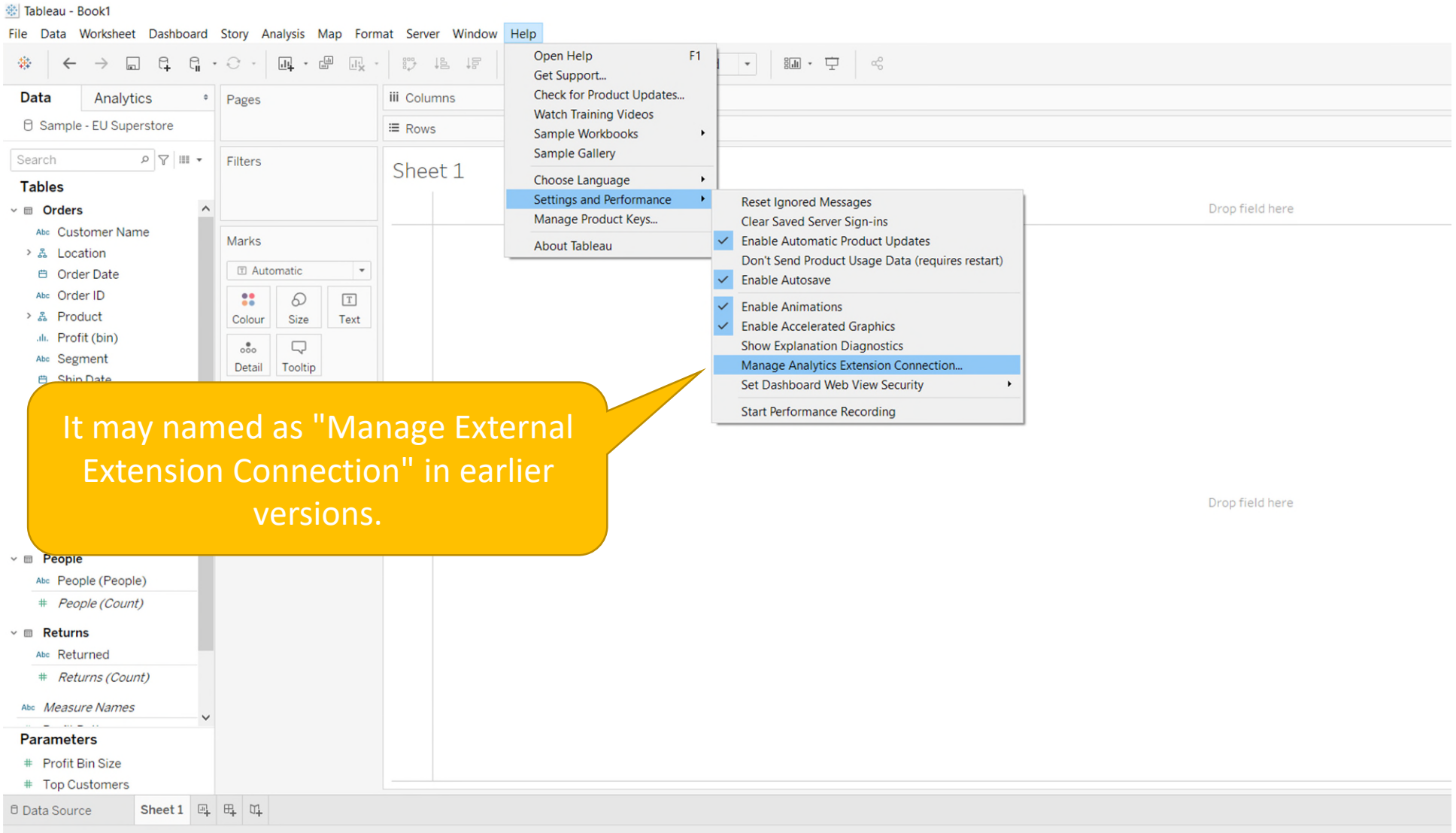


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Open Help F1
Get Support...
Check for Product Updates...
Watch Training Videos
Sample Workbooks
Sample Gallery
Choose Language
Settings and Performance
Manage Product Keys...
About Tableau
Reset Ignored Messages
Clear Saved Server Sign-ins
Enable Automatic Product Updates
Don't Send Product Usage Data (requires restart)
Enable Autosave
Enable Animations
Enable Accelerated Graphics
Show Explanation Diagnostics
Manage Analytics Extension Connection...
Set Dashboard Web View Security
Start Performance Recording
Data
Analytics
Sample - EU Superstore
Search
Tables
Orders
Customer Name
Location
Order Date
Order ID
Product
Profit (bin)
Segment
People
People (People)
People (Count)
Returns
Returned
Returns (Count)
Measure Names
Parameters
Profit Bin Size
Top Customers
Pages
Filters
Marks
Automatic
Colour
Size
Text
Detail
Tooltip
Columns
Rows
Sheet 1
Drop field here
Drop field here
Data Source
Sheet 1
It may named as "Manage External Extension Connection" in earlier versions.

Exercise 1: London Bike Hire

- Santander Cycles is a public bicycle hire scheme in London, Swansea, Milton Keynes and Brunel University in the United Kingdom. The scheme's bicycles used to be popularly known as Boris Bikes, after then-Mayor of London, Boris Johnson. (Wikipedia)
- Please connect Tableau with your Google BigQuery account, add ' bigquery-public-data' project, and find the dataset "london_bicycles". Try to identify potential improvement through data visualization. You may need to write SQL queries to generate the analyzable dataset.
- E.g. stations need more docks, or stations need more bikes.