

Data Connections

1. 4pts Custom Split

Answer using the “Insurance Data” worksheet from the Car Insurance excel file. The identifier field is formatted as a number followed by a “g” (for gender) followed by M or F for male or female followed by “a” for age, followed by a number representing the customer’s age. For example, customer identifier 51gMa16 represents a male age 16. In 2017, how much higher was the average annual insurance premium for males than for females?

- a. \$2.40
- b. \$22.86**
- c. \$16,340
- d. \$56.89

Import the data and use a custom split to determine the age and gender of the customer:

Insurance Data (Car Insurance)

Insurance Data

Sort fields Data source order

Customer Identifier	First Name	Last Name	Email	Gender
1gMa82	Lyon			Male
2gFa22	Fleur			Female
3gFa68	Kiley			Female
4gMa43	Jase	Kubat	jkubat3@ehow.com	Male
5gMa42	Egan	Loisi	eloisi4@wikispaces.c...	Male

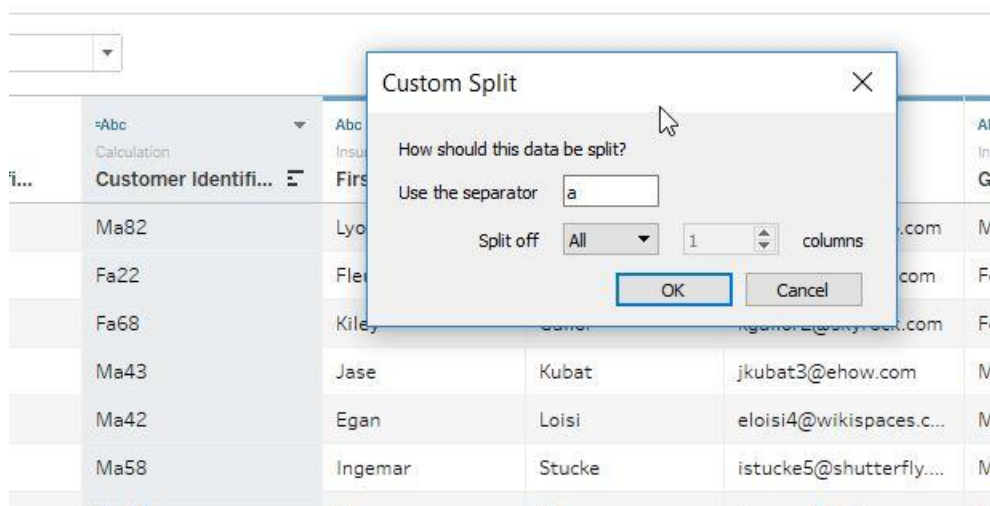
Custom Split

How should this data be split?

Use the separator

Split off columns

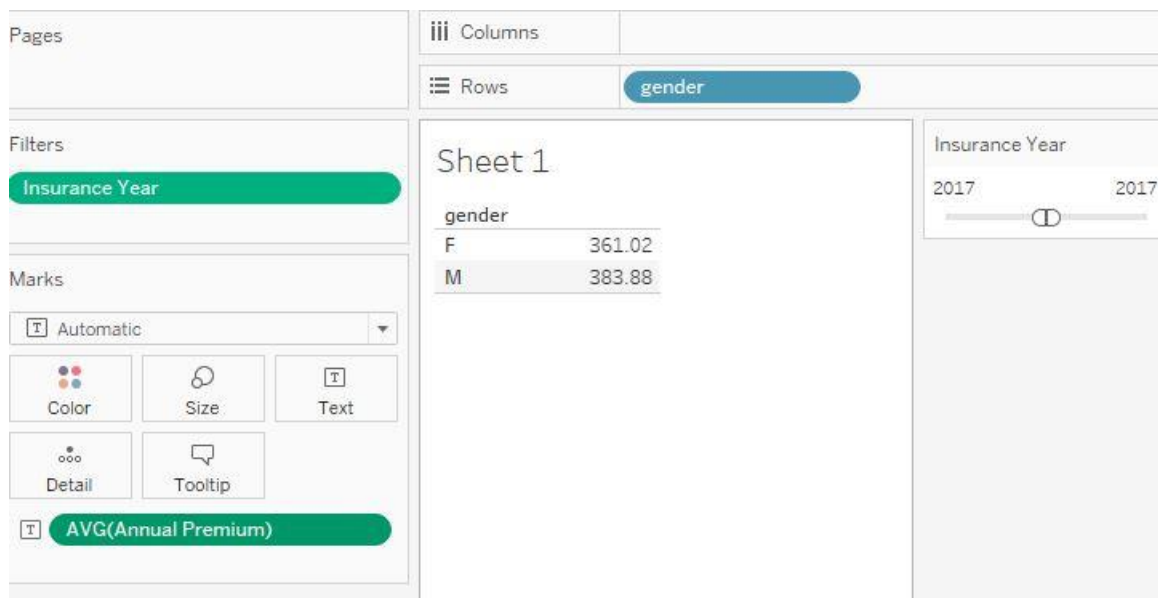
OK Cancel



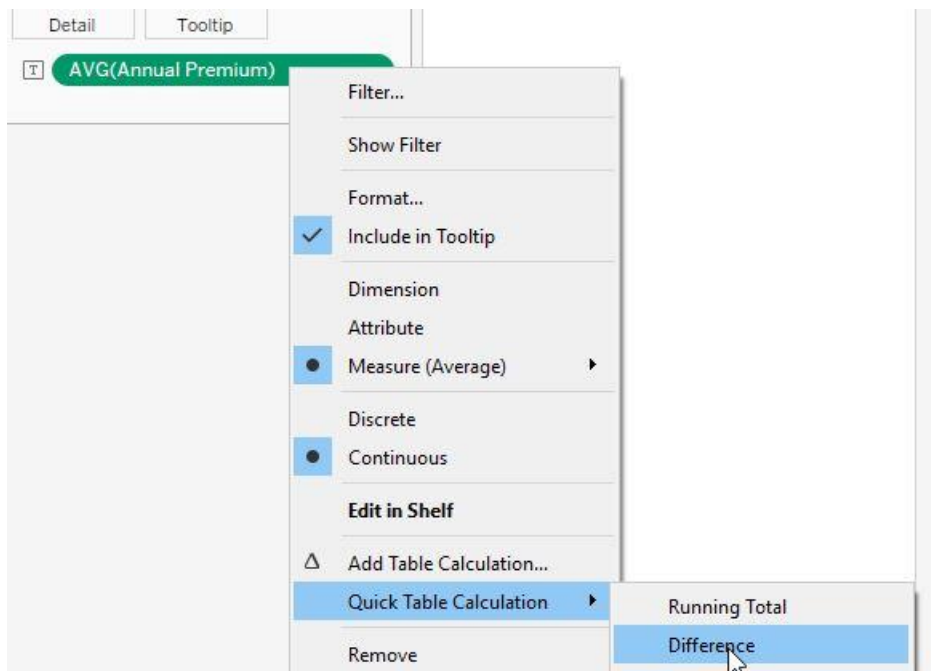
Switch field name to gender

Customer ID	gender	Customer ID
M	82	
F	22	
F	68	
M	43	

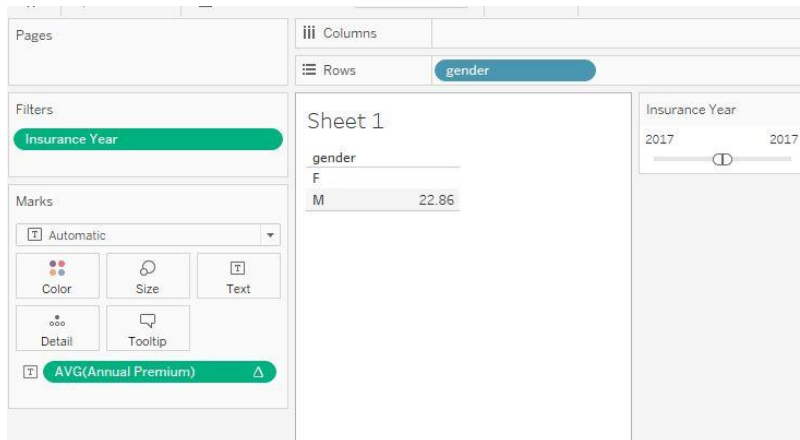
Create view with the average annual premium broken out by gender and the insurance year = 2017



Use a table calculation to determine the difference:



Annual premium for males is 22.86

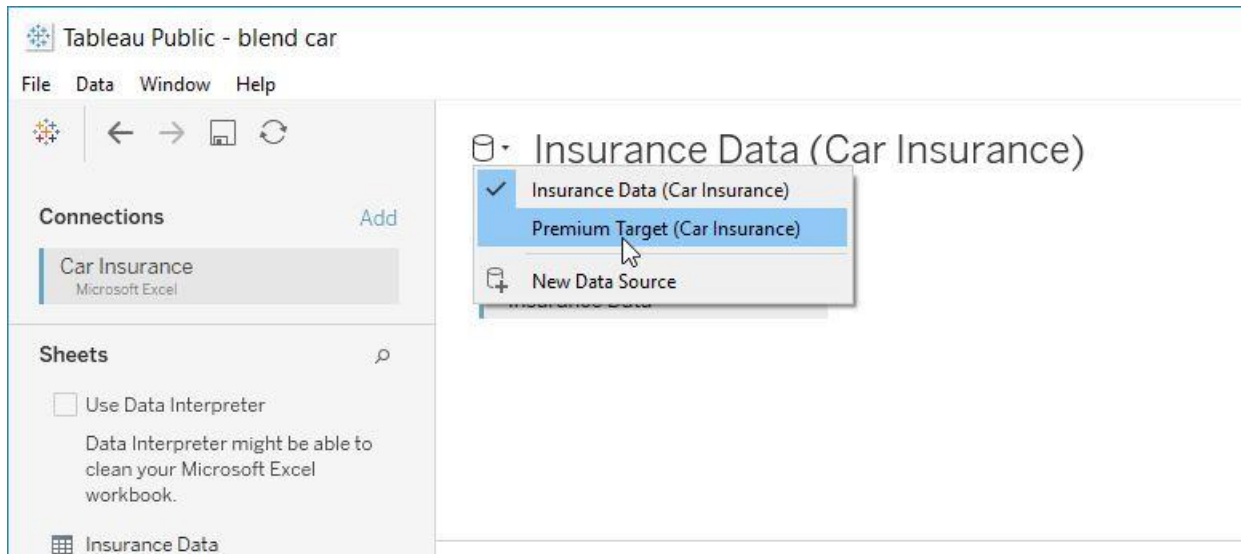


2. 4pts Blend (South Superstore with Target data)

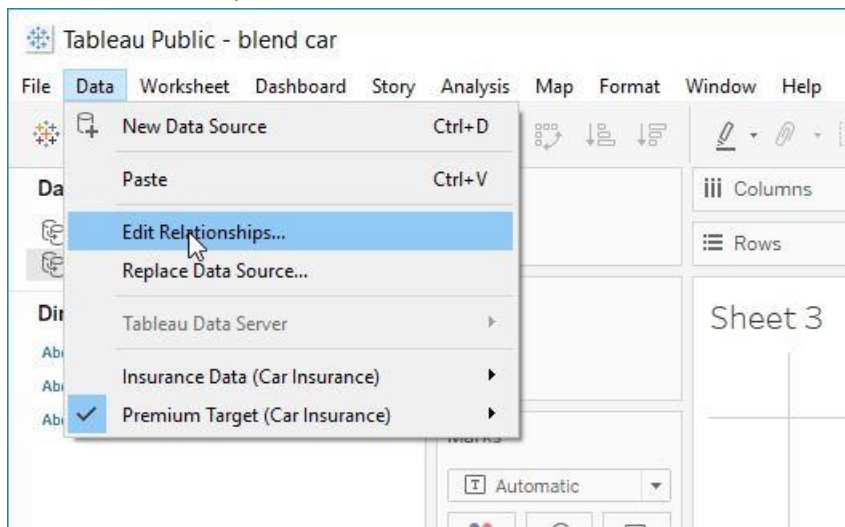
Answer using the "Insurance Data" and "Premium Target" worksheets from the Car Insurance excel file. Which model of Acura had an average annual premium the closest to the targeted premium?

- Legend
- Integra
- RL
- MDX
- ZDX

Create workbook with Insurance Data and Premium Targets:



Edit the relationships for the blend:



Setup the following relationships:

Relationships

Relationships determine how data from secondary data sources are joined with primary data sources.

Primary data source:
Premium Target (Car Insurance)

Secondary data source:
Insurance Data (Car Insurance)

☐ Automatic ☒ Custom

Car Make	Make
Car Model	Model

Add... Edit... Remove

OK Cancel

Add a filter for Make = Acura, add AVG(Annual Premium) and Targeted Premium

Pages

Filters

Measure Names

Make: Acura

Marks

Automatic

Color Size Text

Detail Tooltip

Measure Values

Measure Values

AVG(Annual Premium)

AVG(Targeted Premium)

Columns Measure Names

Rows Model

Sheet 2

Model	Avg. Annual Premium	Avg. Targeted Premium
ZDX	272.3	527.6
TSX	385.8	513.2
Integra	362.3	449.6
NSX	436.0	428.6
RL	382.4	378.9
MDX	375.7	369.9
Legend	321.5	257.5
RDX	451.6	
SLX	372.6	
CL	355.3	
TL	224.9	

Use a calculated field to find the difference between the average annual premium and the target:

Filters

Measure Names

Make: Acura

Marks

Automatic

Color Size Text

Detail Tooltip

Measure Values

AVG(Annual Premium)

AVG(Targeted Premium)

AGG(Target Premium - Premium)

Sheet 2

Model	Avg. Annual Premium	Avg. Targeted Premium	Target Premium - Premium
ZDX	272.3	527.6	255.4
TSX	385.8	513.2	127.3
Integra	362.3	449.6	87.4
NSX	436.0	428.6	-7.3
RL	382.4	378.9	-3.4
MDX	375.7	369.9	-5.7
Legend	321.5	257.5	-64.0
RDX	451.6		0.0
SLX	372.6		0.0
CL	355.3		0.0
TL	224.9		0.0

Target Premium - Premium Insurance Data (Car Insurance)

```
if isnull(SUM([Premium Target (Car Insurance)].[Targeted Premium])) then 0 else
sum([Premium Target (Car Insurance)].[Targeted Premium])-avg([Annual Premium])
end
```

3. [4pts Union] Combine the Insurance Data and the 2019 data. What was the % increase in average annual deductible from 2018 to 2019 for cars made by Toyota?

- a. 2.24%
- b. 4.18%
- c. 3.69%
- d. 3.27%

Do a union of the insurance data (2016 – 2018) and the 2019 data:

Insurance Data+ (Car Insurance)

Insurance Data

Drag table to union

2019 Data

Merge the deductible and 2019 deductible and premium and 2019 premium fields:

☐ Show aliases
 ☐ Show hidden

# Insurance Data+ Deductible	# Insurance Data+ Insurance Year	# Insurance Data+ Annual Premium	# Insurance Data+ 2019 Deductible	# Insurance Data+ 2019 Annual Prem	Abc Insurance Data+ Sheet
null	null	null			3 2019 Data
null	null	null			5 2019 Data
null	null	null			9 2019 Data
null	null	null			2 2019 Data
null	null	null			7 2019 Data
null	null	null	1,780	600.00	2019 Data
null	null	null	760	157.48	2019 Data
null	null	null	2,530	588.93	2019 Data

Rename
 Copy Values
 Hide
 Create Calculated Field...
 Pivot
 Merge Mismatched Fields

“Insurance Year” is null for the 2019 data, so create a calculated field to fill it in:

Insurance Year Corrected

```
ifnull([Insurance Year],2019)
```

Average deductible by year:

The screenshot shows the Tableau Desktop interface. On the left, the 'Marks' card is set to 'Automatic'. Below it, the 'AVG(Deductible & 2019 Deductible)' calculation is displayed. On the right, the 'Columns' shelf contains 'Insurance Year Corre..' and the 'Rows' shelf is empty. The main view shows 'Sheet 1' with a table of data:

Insurance Y..	
2016	1,403.25
2017	1,402.78
2018	1,407.40
2019	1,438.92

Filter on Toyota and use a table calculation to determine the percent difference:

The screenshot shows the 'Add Table Calculation' menu. The 'Quick Table Calculation' option is selected, and the 'Percent Difference' option is highlighted in the submenu.

- Add Table Calculation...
- Quick Table Calculation
 - Running Total
 - Difference
 - Percent Difference
- Remove

Pages

Filters

Make: Toyota

Marks

Automatic

Color

Size

Text

Detail

Tooltip

AVG(Deductible & 2019 Deductible)

Columns

Rows

Insurance Year Corre..

Sheet 1

Insurance Y..	
2016	
2017	-0.039%
2018	0.395%
2019	3.274%

2019 Deductible for Toyota is 3.274% higher than the 2018 deductible.

4. Using the “Average Premium by Make” data, determine which make of car saw the largest cost increase in average premium from 2010 to 2018.
 - a. Land Rover
 - b. Maybach
 - c. Aptera
 - d. Austin

First pivot the data so that instead of having different columns for different cars, there is a “car make” column.

Sort fields

Data source order

ductible By ...	#	#	#	#	#	#	#	#	#	#	#	#
outh	Pontiac	Porsche	Ram	Rolls-Royce	Saab	Saturn	Scion	Shelby	Spyker	Studebaker	Su	
153.867	345.0640	765.443	286.2778	204.674	319.0377	170.262	232.575	170.262				0.063
167.247	328.6323	689.588	260.2525	235.258	370.9741	173.736	250.081	198.736				5.512
209.059	365.1470	638.507	234.4618	245.060	360.1690	197.428	277.868	210.428				3.497
206.989	396.8990	580.461	252.1094	272.289	413.9874	246.785	334.780	230.780				3.445
199.028	367.4990	624.152	283.2690	320.340	376.3521	293.791	348.729	224.729				8.239
248.784	356.7952	685.881	301.3500	299.383	372.6259	330.103	420.156	204.156				9.195
264.664	364.0767	617.911	307.5000	272.167	358.2941	317.406	389.033	200.000	505.000			148.500
371.694	360.6548	404.863	209.4750	314.042	367.7127	415.087	329.292	null	454.091			377.864
403.887	372.5097	322.298	241.2500	487.240	359.6344	389.226	349.189	525.200	546.158			100.901

Rename the fields to “Make” and “Average Annual Premium”

Sort fields: Data source order

#	Abc	#
Avg Deductible By Ma...	Pivot	Pivot
Row Labels	Make	Pivot Field Values
2,010	Acura	263.736
2,011	Acura	280.570
2,012	Acura	264.689
2,013	Acura	294.099
2,014	Acura	358.657

Filter out years other than 2010 and 2018. Add the make and premium to the view, and use a difference table calculation.

Pages: Columns: Measure Names Rows: Make Insurance Year

Filters: Insurance Year Measure Names

Marks: Automatic Color Size Text Detail Tooltip Measure Values SUM(Average Annual Premium) SUM(Average Annual Premium)

Sheet 1

Make	Insurance Y..	Average Annual Premium	Difference in Average Annual Premium from the Pr..
Acura	2010	355	
Acura	2018	401	46
Alfa Romeo	2010	514	
Alfa Romeo	2018	464	-50
Aptera	2010	147	
Aptera	2018	534	386
Aston Martin	2010	371	
Aston Martin	2018	362	-10
Audi	2010	163	
Audi	2018	333	171
Austin	2010	412	
Austin	2018	742	330
Bentley	2010	271	
Bentley	2018	329	58
BMW	2010	318	
BMW	2018	402	84

Insurance Year: (All) 2010 2011 2012 2013 2014 2015 2016 2017 2018

Aptera’s premium increased by 386:

Make	Insurance Y..	Average Annual Premium	Difference in Average Annual Premium from the Pr..
Acura	2010	355	
Acura	2018	401	46
Alfa Romeo	2010	514	
Alfa Romeo	2018	464	-50
Aptera	2010	147	
Aptera	2018	534	386
Aston Martin	2010	371	
Aston Martin	2018	362	-10
Audi	2010	163	
Audi	2018	333	171

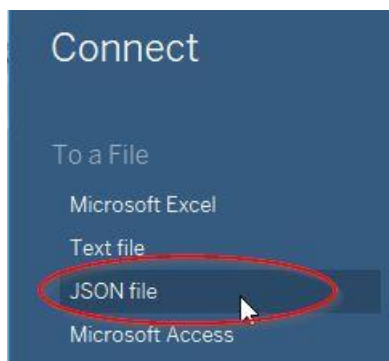
Insurance Year: 2018
Make: Aptera
Difference in Average Annual Premium from the Previous along Pane (Down): 386

5. 4pts Connect to different file type (JSON, CSV, etc)

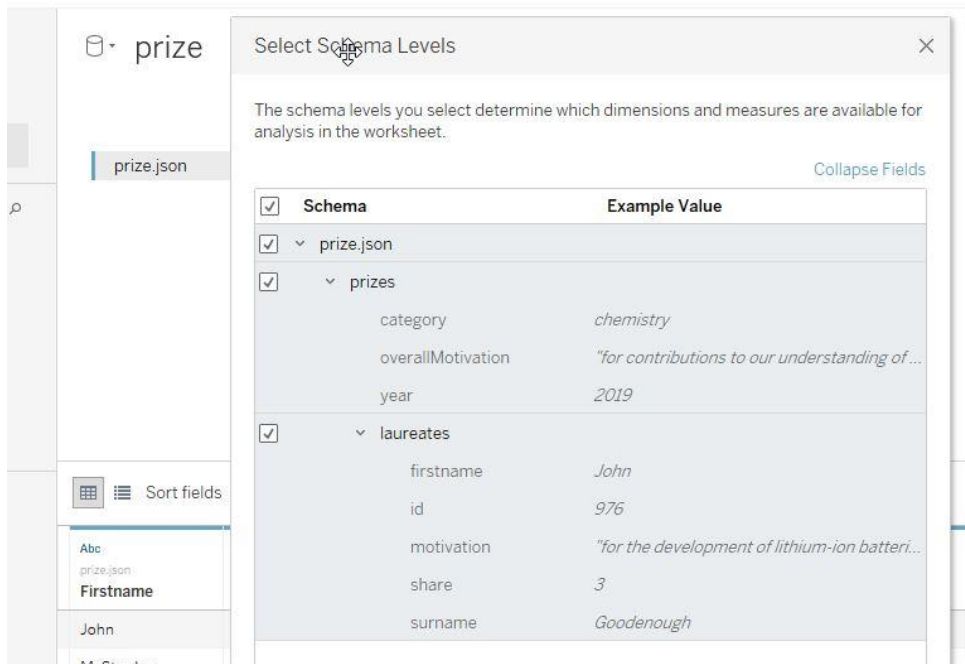
The Nobel Prize is a set of annual international awards bestowed in several categories. Sometime years the prize is awarded to a single individual or organization, other times it is shared, and other times there is no award in a category in a particular year. Using “nobel_prize.json,” for the years when the prize is awarded in a category, what percent of the time is it shared?

- a. 41.37%
- b. 58.63%
- c. 45.82%
- d. 54.18%

Connect to the JSON file



Select the schema



Create a view to understand the data:

Year	Category	Firstname	Surname	Share	Row
1965	medicine	André	Lwoff	3	Abc
		François	Jacob	3	Abc
		Jacques	Monod	3	Abc
	peace	United Nations Children's ...	Null	1	Abc
	physics	Julian	Schwinger	3	Abc
		Richard P.	Feynman	3	Abc
		Shin-Itiro	Tomonaga	3	Abc
1966	chemistry	Robert S.	Mulliken	1	Abc
	literature	Nelly	Sachs	2	Abc
		Shmuel	Agnon	2	Abc
	medicine	Charles B.	Huggins	2	Abc
		Peyton	Rous	2	Abc
	peace	Null	Null	Null	Abc
1967	physics	Alfred	Kastler	1	Abc
		George	Porter	4	Abc
		Manfred	Eigen	2	Abc
	chemistry	Ronald G.W.	Norrish	4	Abc
		Miguel Ángel	Asturias	1	Abc
		George	Wald	3	Abc
	medicine	Keffer	Hartline	3	Abc
		Ragnar	Granit	3	Abc
1967	peace	Null	Null	Null	Abc
	physics	Hans	Bethe	1	Abc

Notice that in 1966 and 1967 the nobel prize was not awarded to any person(s). Since the question asks, "When the nobel prize is awarded in a category..." we want to eliminate the times when the nobel prize *is not* awarded in any category. This can be done by removing the rows where [Share] is null:

Filter [Share]

General
Wildcard
Condition
Top

☒ Select from list
☐ Custom value list
☐ Use all

Enter search text

☐ Null
☒ 1
☒ 2
☒ 3
☒ 4

All
None
☐ Exclude

Summary
Field: [Share]
Selection: Selected 4 of 5 values
Wildcard: All
Condition: None
Limit: None

Reset
OK
Cancel
Apply

Now also we see that when [Share] = 1, the prize is awarded only to one person in that category in that year.

Total prizes awarded:

Prizes awarded

```
countd([Year]+[Category])
```

Prizes awarded to just one person:

Prizes awarded to 1 person

```
countd(if [Share]='1' then [Year]+[Category] else null end)|
```

Result:

Sheet 1	
Prizes awarded to 1 person	350.0
Prizes awarded	597.0

Share

☐ (All)

☐ Null

☒ 1

☒ 2

☒ 3

☒ 4

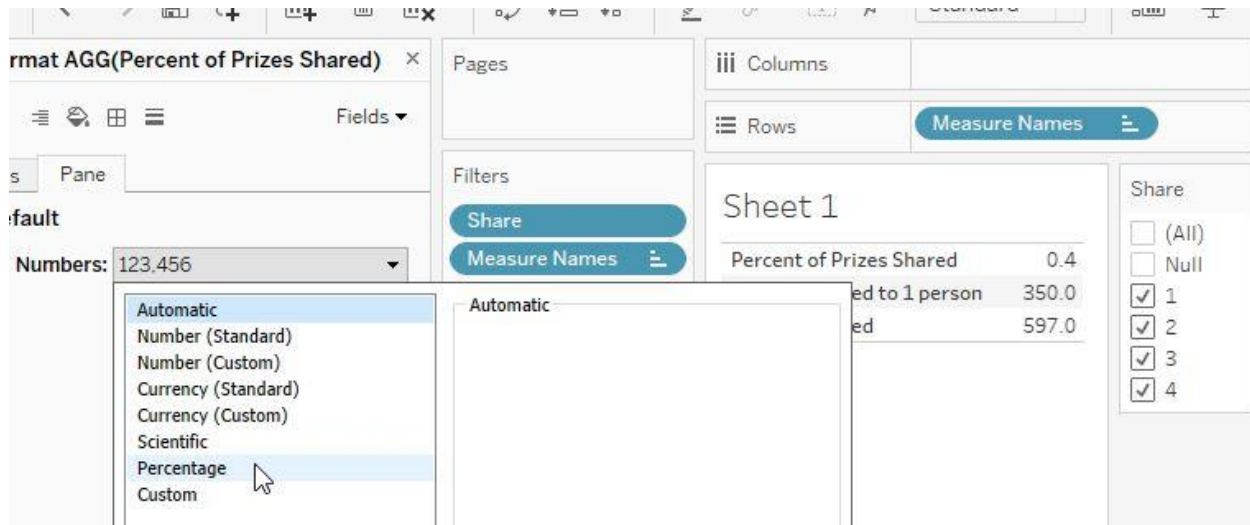
$350 / 597 = 58.6\%$ (rounded) of prizes were awarded to just one person, so around 41.4% are awarded to more than one person. To do that in Tableau, we can use a calculation like this:

Percent of Prizes Shared

```
countd(if [Share] = '1' then null else [Year] + [Category] end)
/
countd([Year] + [Category])
```

The numerator is the distinct prize years awarded to multiple people (shared). Since we've filtered out the cases where the prize was not awarded at all, the denominator is the total prizes awarded.

Add this to the view and change the format to %



Result:

Sheet 1	
Percent of Prizes Shared	41.37%
Prizes awarded to 1 person	350.0
Prizes awarded	597.0

6. 2pts Knowledge – Performance Optimization Which of the following can reduce the number of queries Tableau needs to run to load a worksheet or dashboard
- Parallel query
 - Data engine vectorization
 - Data source filters
 - External query caching
 - Metadata grid

Parallel Queries - Tableau 9.0 also takes better advantage of the capabilities of source databases to execute more queries at the same time. This new "Parallel Queries" feature will improve Dashboard performance, on Desktop and Server. <https://www.tableau.com/about/blog/2015/1/90-preview-query-performance-improvements-36406>

Data engine vectorization – performs operations such as plus, minus, divide, min, max, sum, etc. on multiple data in parallel, but does not reduce the number of queries
<https://community.tableau.com/thread/211729>

Data source filters – can reduce the rows returned by the query, but not the number of queries.
https://help.tableau.com/current/pro/desktop/en-us/filtering_datasource.htm

External query caching

When loading a workbook for the first time, Tableau queries the data source to get the values to create the visualizations. However, what happens the second time? If the data hasn't changed, like with extract-based workbooks, why even query the data source?

Tableau 9.0 will save the query cache as part of the workbook on both Desktop and Server, offering near-instant load times for applicable workbooks. There is no need to re-query for answers since the results are already there, and users can simply refresh to replace the cache.

Imagine simply re-opening a workbook that has been opened locally before, or that has already been viewed on Server, and seeing a virtually instant load time.

<https://www.tableau.com/about/blog/2015/1/90-preview-query-performance-improvements-36406>

Metadata grid - The metadata grid displays the fields in your data source as rows so that you can quickly examine the structure of your Tableau data source and perform routine management tasks, such as renaming fields or hiding multiple fields at once. Does not reduce the number of queries.

7. 1pts Knowledge – Shadow Extracts – Which of the following is true of shadow extracts?
- a. The shadow extract file extension is .hyper or .tde
 - b. Shadow extracts can be created by changing the connection type from “Live” to “Extract” on the data source page.
 - c. Shadow extracts are created automatically when certain data source connections are created.
 - d. You can choose to create a shadow extract with all rows, top N rows, or a sample of the rows in the underlying data.

Shadow extracts use .hhyper and .ttde extension rather than .hyper and .tde

Extracts are created by switching from “Live” to “Extract” on the data source page

https://help.tableau.com/current/pro/desktop/en-us/extracting_data.htm#create-an-extract

See post <https://community.tableau.com/thread/201115>

When you work with Tableau Desktop 9.0 and later, your My Tableau Repository might contain a folder called Shadow Extracts. By default, the Shadow Extracts folder can contain up to five shadow extract files, which have a .ttde extension.

The contents of the Shadow Extracts folder change depending on the workbook you work with, and shadow extract files are only created when you work with workbooks that are based on non-legacy Excel or text, or statistical files.

Tableau creates and saves a shadow extract file in order to load your data more quickly. After Tableau creates five shadow extract files, Tableau deletes the oldest shadow extract file to create space when it adds a new one.

Although shadow extract files contain underlying data and other information similar to the standard Tableau extract, shadow extract files are saved in a different format, which means that they cannot be used the same way Tableau extracts are.

Calculations

8. [manipulate string and date calculations] Answer the following question using the Southern Region sheet of the South Superstore data. Determine the number of days from the order date to the ship date for each Row ID. Which of the following has the largest average difference in shipping time?
- Same day vs First Class
 - First class vs Second Class
 - Second Class vs Standard Class
 - Same day vs Freight Class

Create the days to ship calculation

Days to Ship

`DATEDIFF("day",[Order Date],[Ship Date])`

The calculation is valid.
1 Dependency
Apply
OK

All
Enter search text
COS
COT
COUNT
COUNTD
COVAR
COVARP
DATE
DATEADD
DATEDIFF

DATEDIFF(date_part, start_date, end_date, [start_of_week])

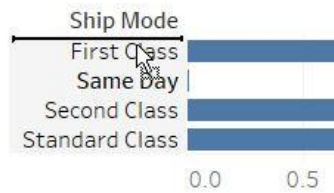
Returns the difference between two dates where start_date is subtracted from end_date. The difference is expressed in units of date_part. If start_of_week is

Create a bar graph showing days to ship:

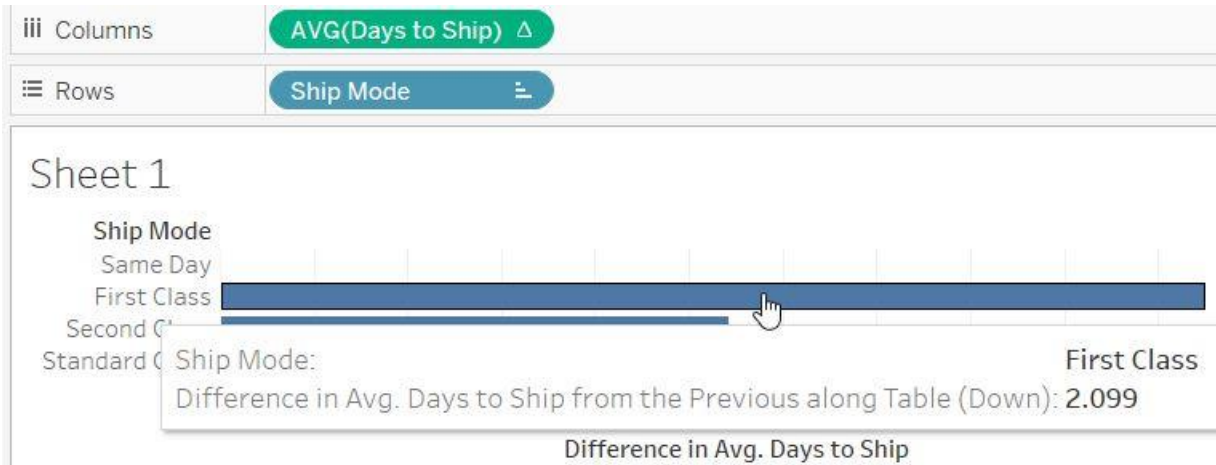


Reorder the ship modes so that same day is at the top:

Sheet 1



Use a difference table calculation to determine the difference between same day and first class, first class and second class, and second class and standard class.



The days to ship for first class is 2.099 higher than for same day, larger difference than any of the other options. Freight Class does not exist in the data set.

9. [LOD calculations] Answer the following question using the Southern Region sheet of the South Superstore data. For customer's making their first purchase in 2014, how many placed three or more separate orders?
- 67
 - 51
 - 23
 - 26

Unique purchases by customer using Fixed LOD

Purchases By Customer

```
{fixed [Customer ID]:countd([Order ID])}
```

Customer's first order year

Customer's 1st Order Year

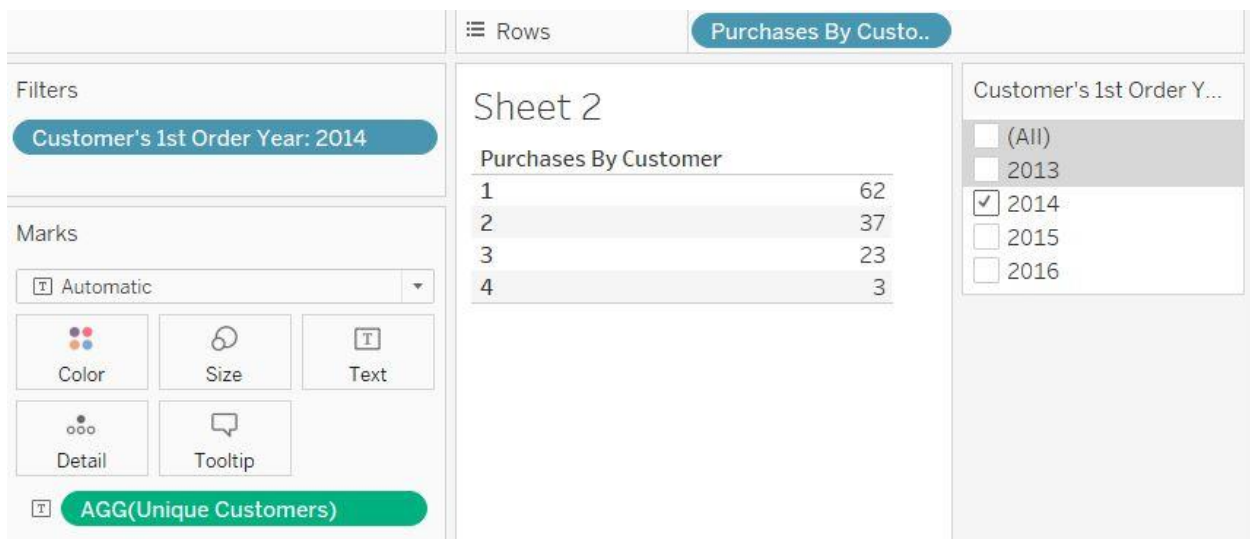
```
{fixed [Customer ID]:min([Order Year])}
```

Unique Customers:

Unique Customers

```
countd([Customer ID])
```

This view shows purchases by customer. 23 customers made 3 purchases, and an additional 4 customers made 4 purchases, so 26 customers made 3 or more purchases.



10. [Calculations in join clauses] During the 2018 insurance year, individuals affiliated with some organizations were able to get an insurance premium discount. The premium value shown in "Insurance Data" does not include the discount. The "Domain Discount" shows the discount amount based on the second half of the customer's email address... for example, Lukas@nasa.gov would use the nasa.gov domain. Find the domain with the highest total discount amount.
- squarespace.com
 - ning.com
 - dropbox.com
 - meetup.com

Drag "User Data" and "Domain Discount" into the data source area and create a join using a calculated field:

Join calculation

```
split([Email], "@", 2)
```

The calculation is valid.

Apply
OK

Insurance Data+ (Car Insurance)

Insurance Data
Domain Discount

Join

Inner
Left
Right
Full Outer

Data Source		Domain Discount
split([Email], "@", 2)	=	Domain
Add new join clause		

Calculate the discount amount for each row:

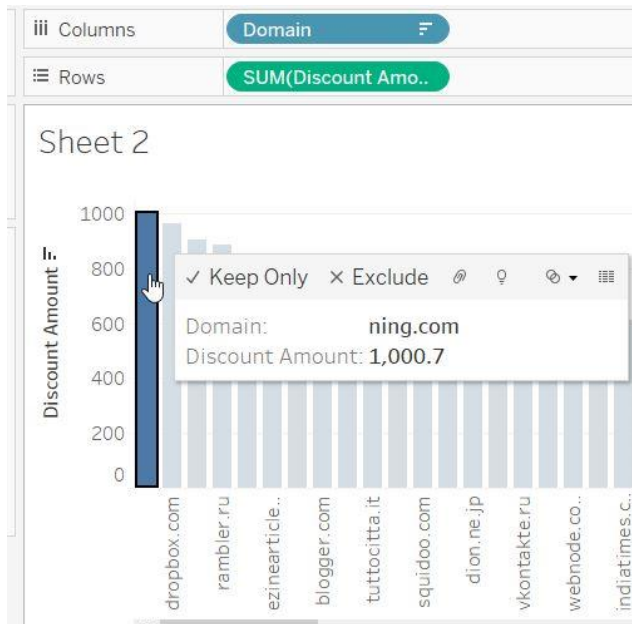
Discount Amount

```
[Annual Premium] * [Premium Discount]
```

The calculation is valid.

Apply
OK

Determine the discount amount:



11. [3pts Build logic statements] Using the Southern Region data from the South Superstore worksheet, find the number of unique records with negative profit.

a. 259

b. 272

c. 1,361

d. 1,389

e. 1,620

unique rows

countd([Row ID])

Profit Negative

```
if [Profit]<0 then 'negative'
else 'positive or zero'
end
```

Pages	Columns	
Filters	Rows	Profit Negative
Marks	Sheet 1	
Automatic	Profit Negative	
Color	negative 259	
Size	positive or zero 1,361	
Text		
Detail		
Tooltip		
AGG(unique rows)		

12. 1pts Knowledge: types of LOD calculations – The expression {SUM([Profit])} will return which of the following:

- a. An error
- b. The total profit for the entire table
- c. The profit for each row individually
- d. The profit for the current dimensions in the view

This is a table-scoped level of detail expression [as shown here](#). This is equivalent to {FIXED:SUM([Profit])}. It will return the sum of Profit for the entire table.

13. 1pts Knowledge: totals or subtotals – Which of the following “Total all using” options cause the total and subtotal calculations to be based on the underlying data rather than the data in the view? (select all that apply)

- a. Automatic
- b. Sum
- c. Average
- d. None of the above

When you choose Automatic, totals are based on the underlying data, which is disaggregated, and not on the data in the view. When you choose any of the other values (Sum, Average, Minimum, or Maximum), all totals are computed using the selected aggregation. The computations are performed on the aggregated data you see in the view. https://help.tableau.com/current/pro/desktop/en-us/calculations_totals_grandtotal_turnon.htm#Configure

Fields & Chart Types

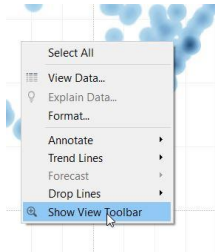
14. [4pts Heat Map] Using the Southern Region sheet of the South Superstore data, create a density heat map showing disaggregated sales and profit. Which customer segments show a high density area where Profit is 6 and Sales is 13? [you may select multiple answers]

- a. Consumer
- b. Corporate
- c. Home Office

d. None of the Above

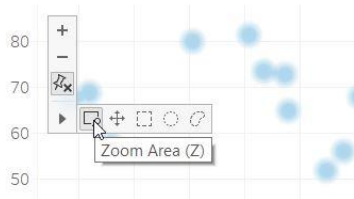
Create a scatter plot with Profit and Sales. Uncheck aggregate measures. Change the marks type to density. Add a filter for segment.

Right click and show toolbar:



Zoom in on the relevant area of the graph:

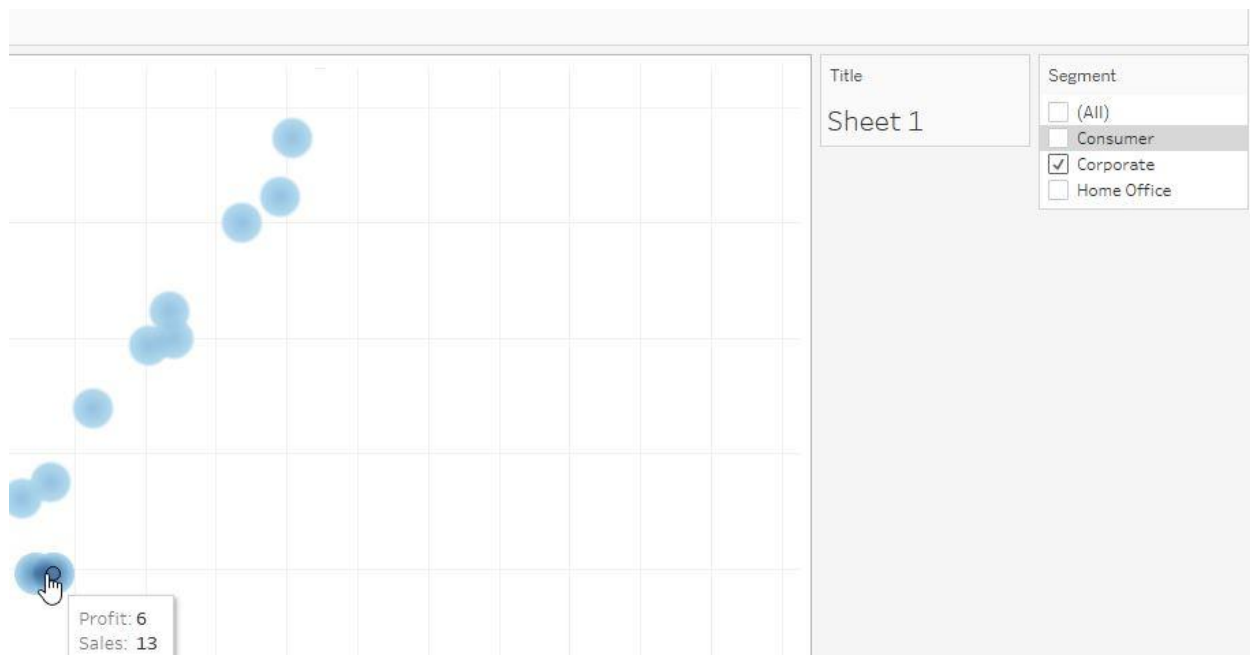
Sheet 3



Consumer



Corporate:



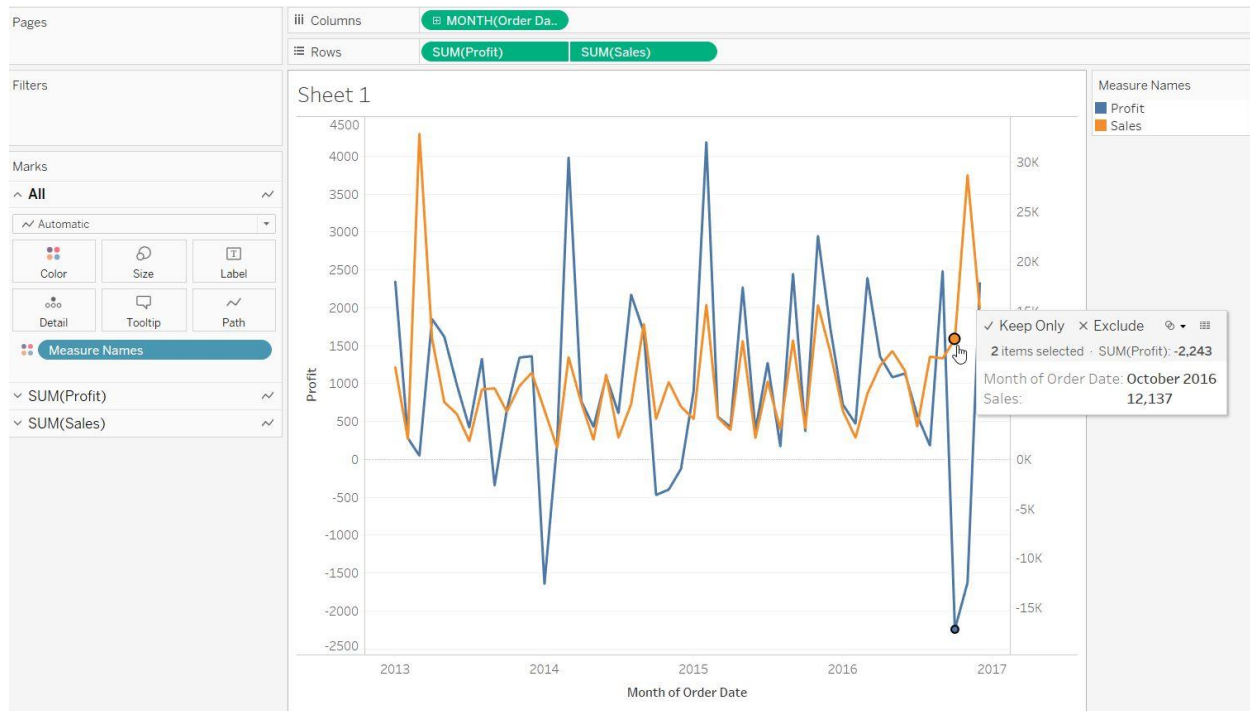
Home office:



15. [4pts Dual Axis Chart] Using the Southern Region sheet of the South Superstore data, plot Sales and Profit by month. Profit and Sales tend to increase and decrease together. In what month is does sales increase while profit decreases?

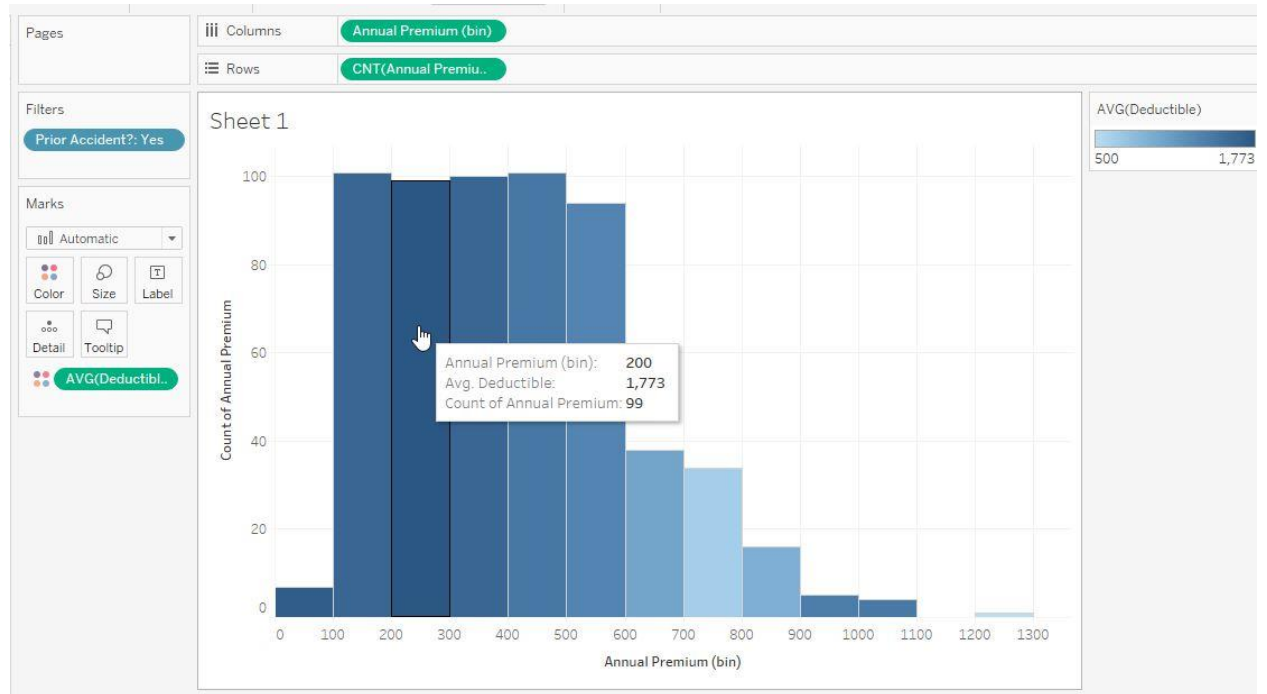
- a. October 2014
- b. November 2104
- c. October 2016
- d. November 2016

Profit decreased while sales increased in October 2016



16. [4pt Histogram Chart] Using the Insurance Data worksheet from the Car Insurance workbook, create a histogram showing the number of insurance policies using Annual Premium bins of \$100 for just those policies where the customer had a prior accident. Which bin has an average deductible about \$1,800?

- a. 100
- b. 200**
- c. 300
- d. 400
- e. 500



17. 1pts Knowledge: Understand discrete v. continuous – Which of the following will typically be discrete?

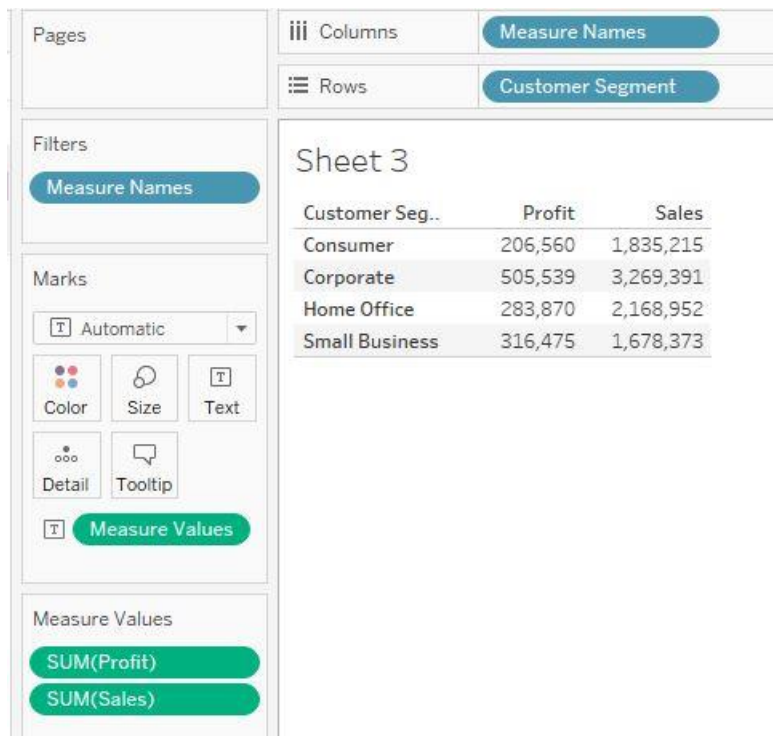
- a. Height in centimeters
- b. Employee ID**
- c. Price
- d. Speed
- e. None of the above

All of these fields are numeric, but employee ID is usually treated as a discrete dimension by default because there is no continuum of values for employee IDs... 70 centimeters in height is less than 71 centimeters in height, but employee ID 1743 is not less than the employee ID 1744.

18. 1pts Knowledge: Understand measure names and measure values – You would like a text table with one column showing sum of Sales and another column showing sum of profit. What pill should be in the column shelf?

- a. Measure Names**
- b. Measure Values
- c. Nothing – the column shelf should be empty for this view
- d. SUM([Sales]) and SUM([Profit])

Here we see that the “Measure Names” pill is in the Columns shelf



19. 1pts Knowledge: Understand chart types – Which of the following chart types uses a running total?

- a. Sparklines
- b. Pareto**
- c. Gantt charts
- d. Bullet Graphs
- e. None of the above

A Pareto chart is a type of chart that contains both bars and a line graph, where individual values are represented in descending order by bars, and the ascending cumulative total is represented by the line.

<https://help.tableau.com/current/pro/desktop/en-us/pareto.htm>

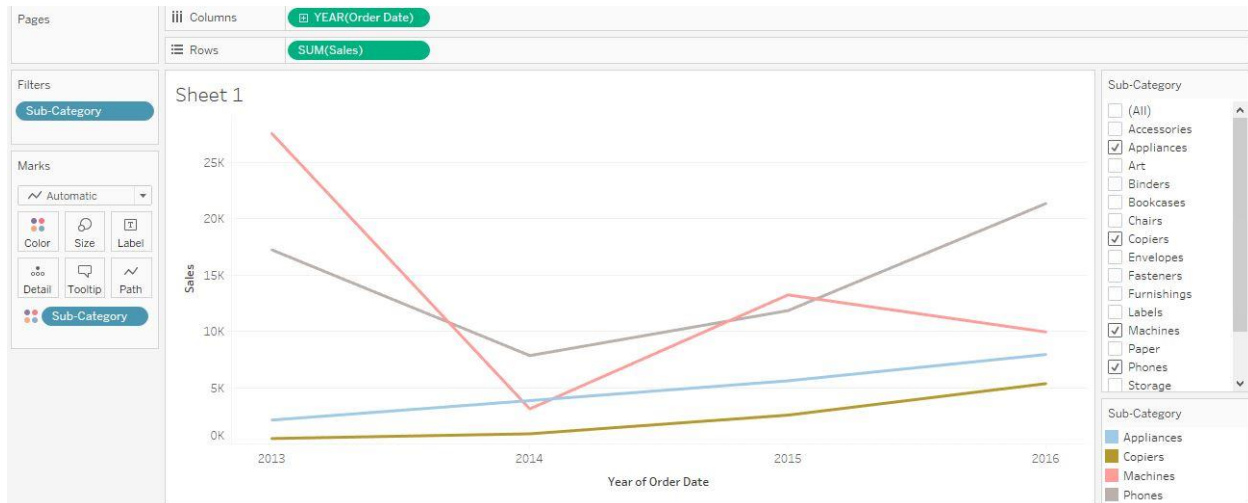
Analytics

20. [4pts – Trend Line] Using the Southern Region worksheet of the South Superstore Data, create a linear trend line showing the relationship between sales and year of order date for each product subcategory. Which of the following subcategory has seen the greatest average increase in sales?

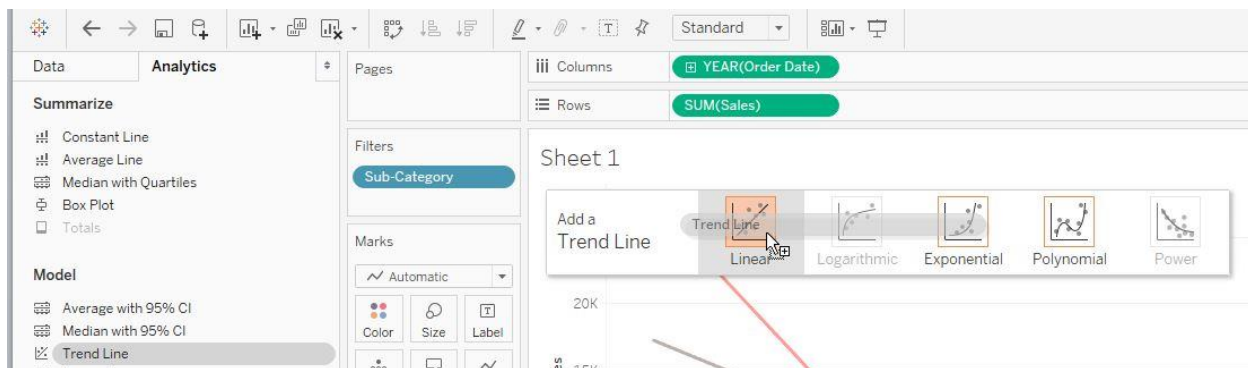
- a. Appliances**
- b. Phones
- c. Copiers
- d. Machines

Set a filter on just the sub-categories, so that only the four listed in the question are selected.

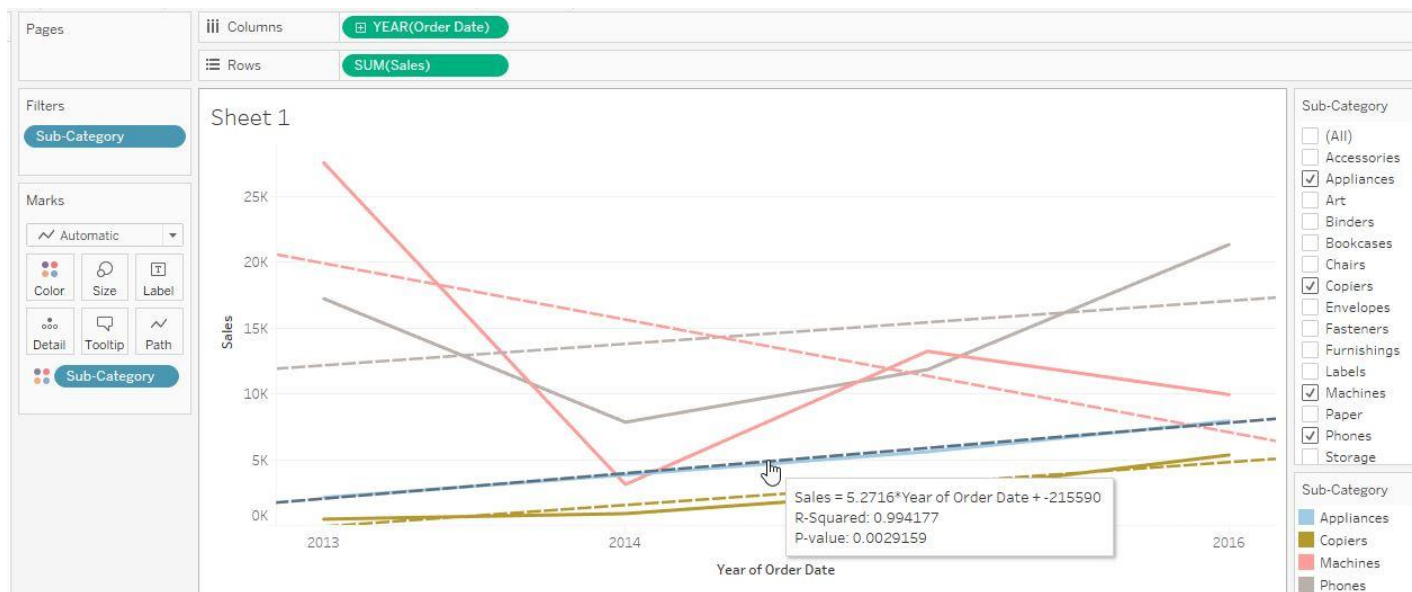
Then create a line graph



Next add a linear trendline:



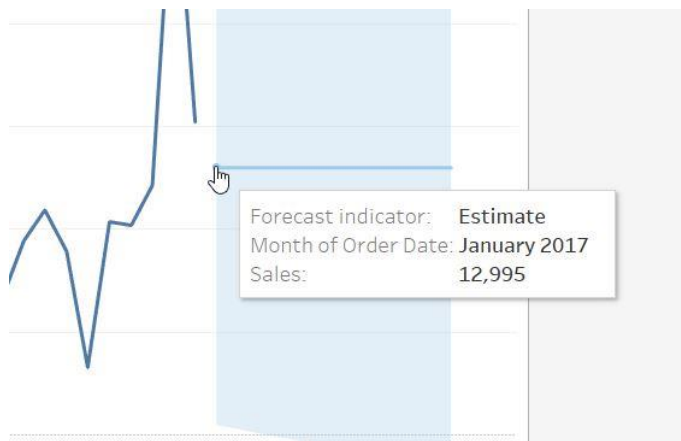
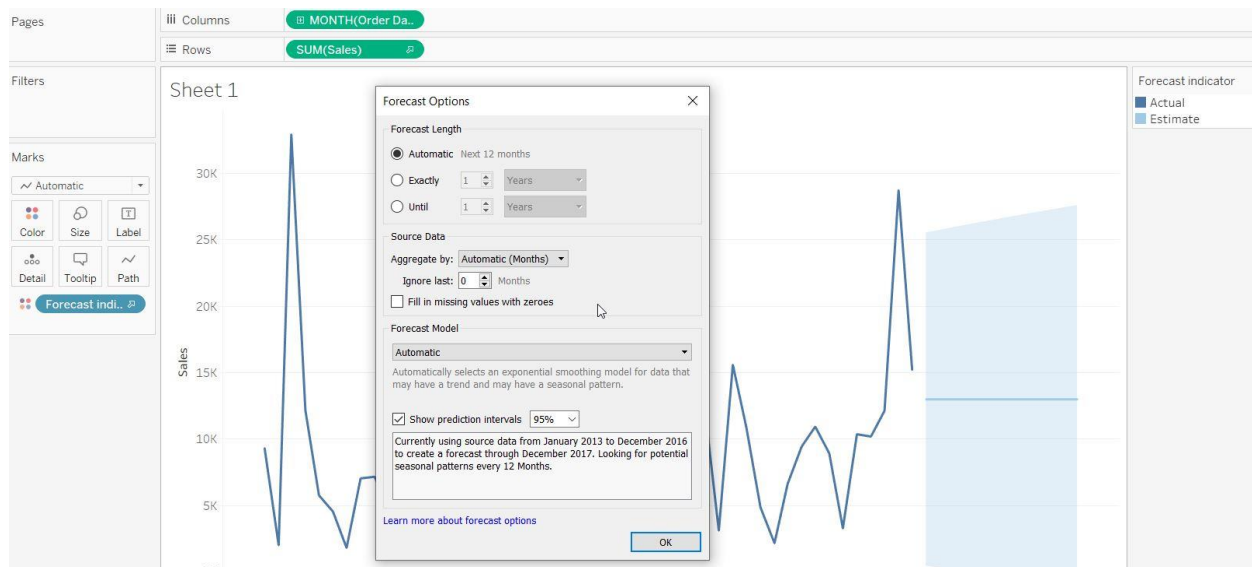
Mouse over the four trendlines. The coefficient of Year of Order Date is the average increase in sales per year. The value is highest for appliances, indicating the highest average increase per year for that subcategory.



21. [4pts – Forecast] Using the Southern Region worksheet of the South Superstore Data, forecast the sales total for January 2017 using the monthly sales totals through December 2016.

- 10,724
- 12,995**
- 12,509
- 15,210

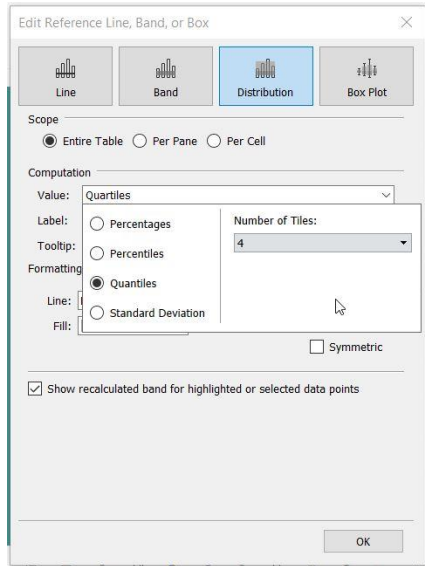
Plot sales by month of order date. Right click and select forecast options, and select Ignore Last 0 Months so that the forecast uses the sales totals through December 2016.



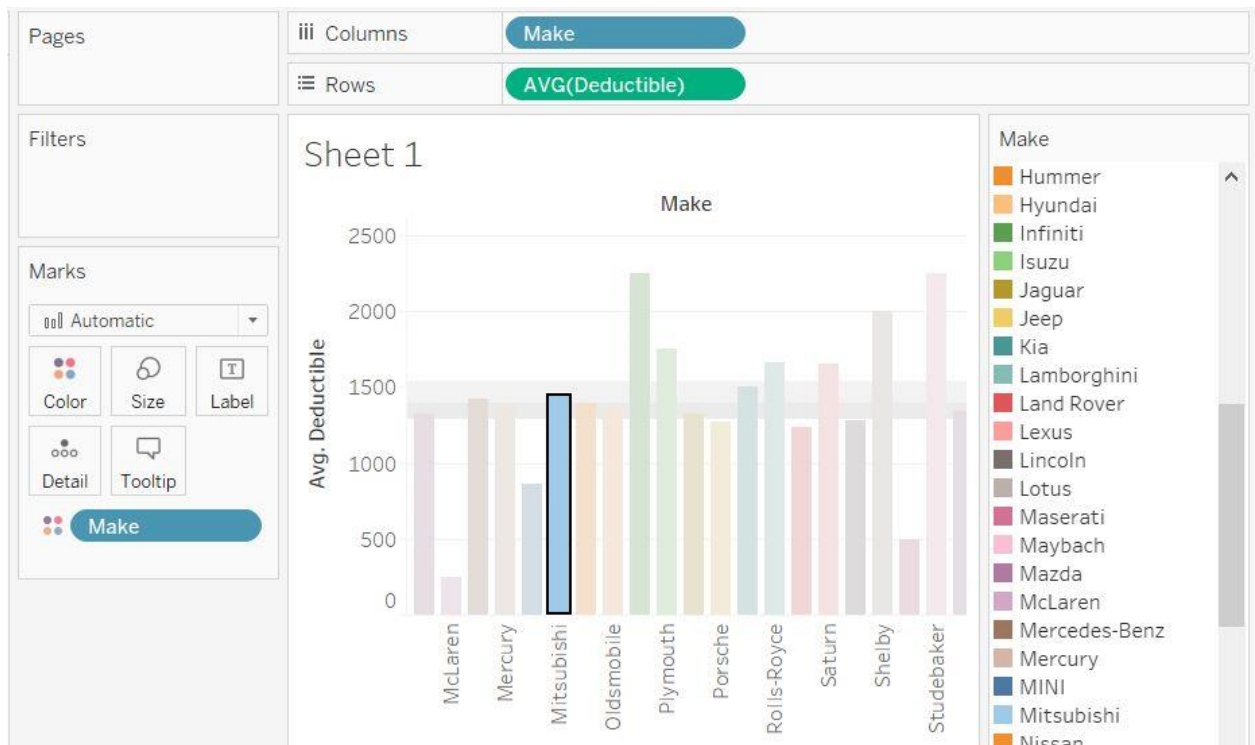
22. [4pts – Compare median and mean] Using the Insurance data, create a chart showing the average deductible by make of car. Which model has an average deductible between the median and the 75th percentile?

- Toyota
- Mitsubishi
- Bentley
- Peugeot

Create a bar chart for avg(deductible) by model. Add a reference distribution with 4 tiles:



This shows the Mitsubishi is between the median and the 75th percentile:

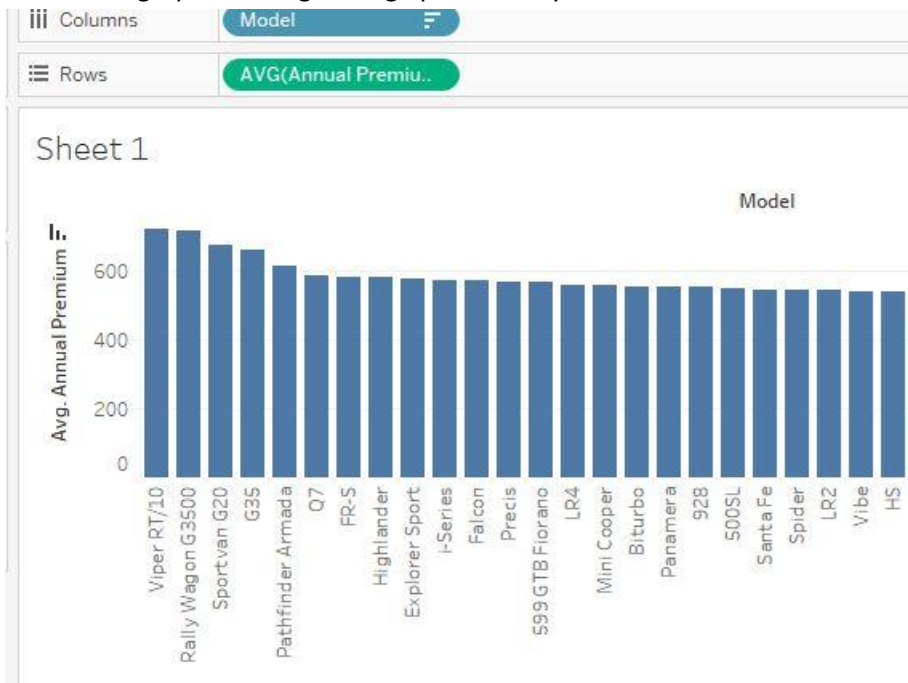


23. 2pts – Find skewness (using statistical summary card) – Using the “Insurance Data” worksheet from the “Car Insurance Workbook. Find the skewness for the distribution of average annual premium costs by car Model.
- 2.04
 - 2.09
 - 0.50
 - 0.11

e. 0.01

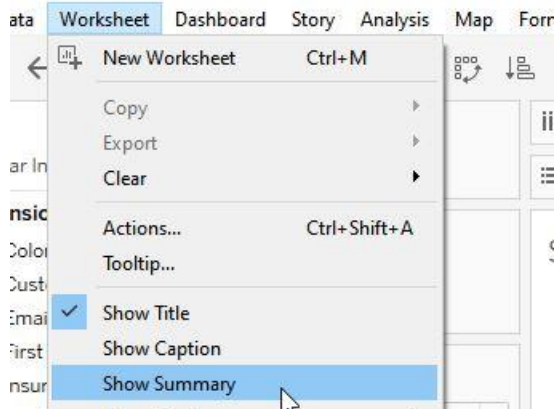
Documentation here: https://onlinehelp.tableau.com/current/pro/desktop/en-us/inspectdata_summary.html

Create bar graph showing average premium by model:

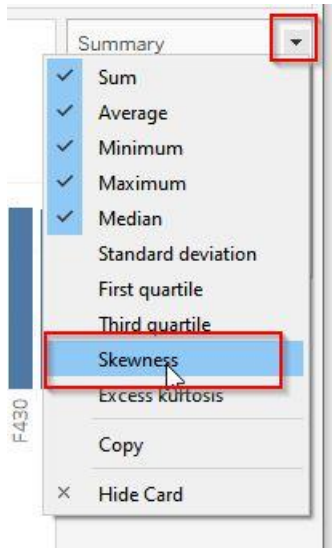


Select "Show Summary" to show statistical summary:

Tableau Public - Book3



Click arrow on summary card and select Skewness:



View should now shows 0.50 skewness:



24. 1pts – Which part of a box plot indicates the interquartile range?

- Upper hinge to lower hinge
- Median to upper hinge
- Upper whisker to lower whisker
- Upper whisker to median
- Median to lower hinge

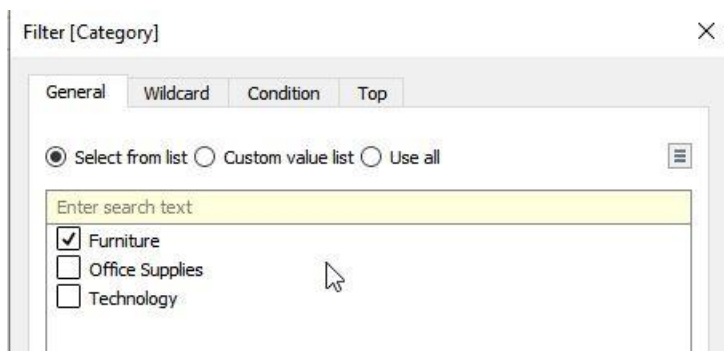
https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_boxplot.html

Organizing and Simplifying Data

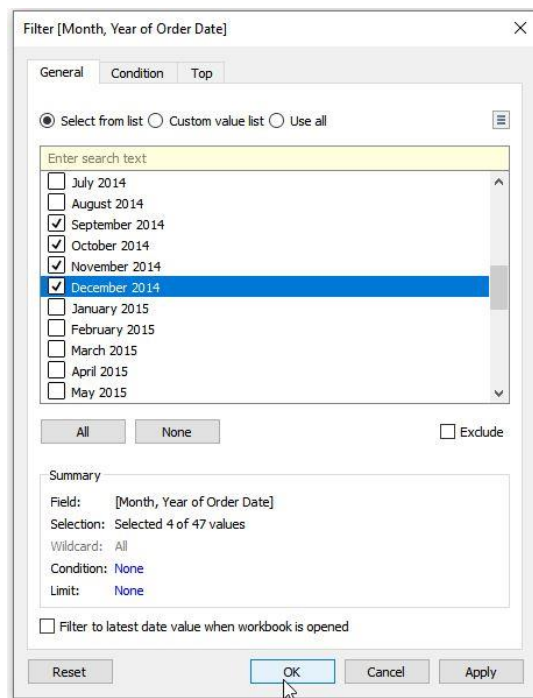
25. [4pts Filter data] Using the Southern Region worksheet of the South Superstore data, find the average number of unique orders which included furniture per month for the last four months of 2014.

- a. 9.67
- b. 12
- c. 20.33
- d. 8.00
- e. None of the above

Filter on Furniture:



Filter on Order Date in the last 4 months of 2014:

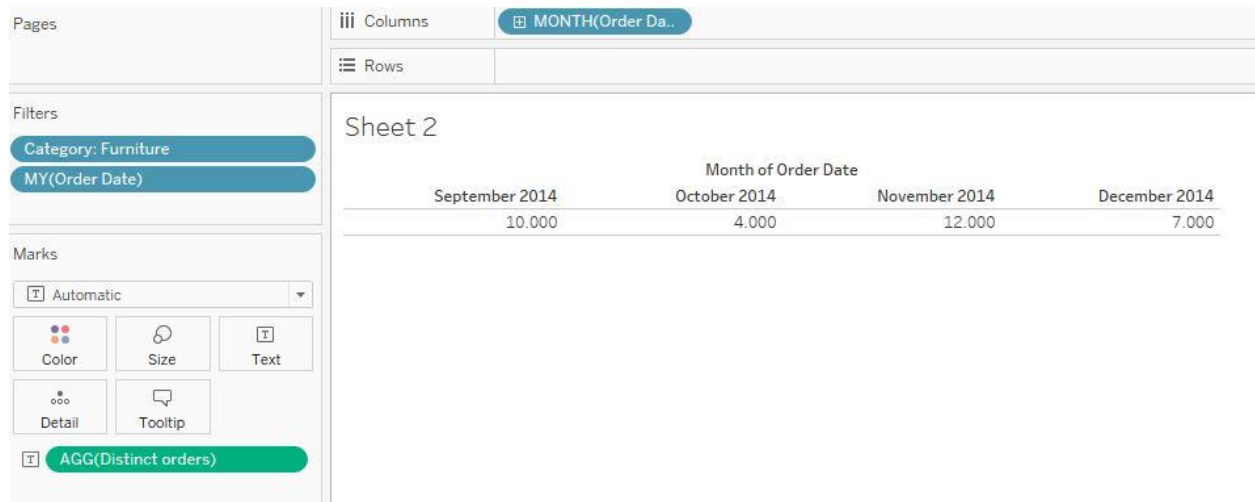


Distinct orders calculation:

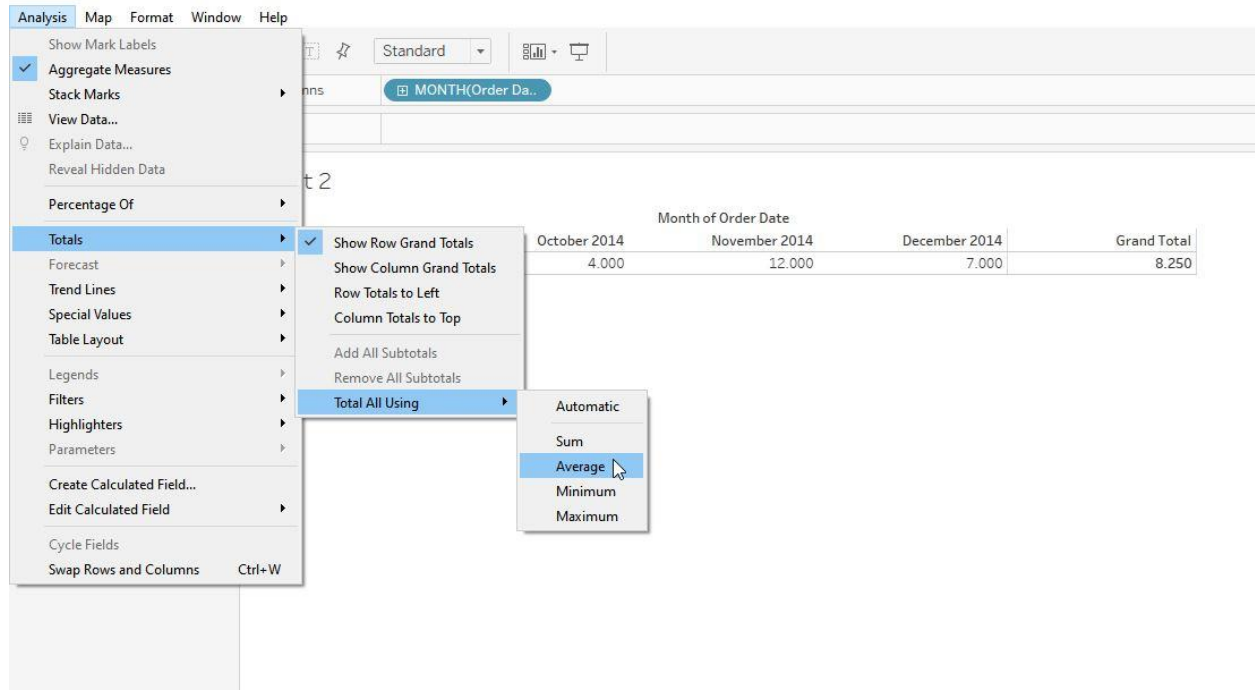
Distinct orders

```
countd([Order ID])
```

View:



Monthly average will be $(10 + 4 + 12 + 7) / 4$. We can do this using grand total with total using average:



The average monthly orders including furniture for October to December 2014 is 8.25, so the correct answer is none of the above.

26. [4pts Build sets] Answer using the “Insurance Data” worksheet from the Car Insurance excel file. For cars insured in New York state, find the top 10 to makes of car with the highest average annual premium. Of these 10, which had the most policies associated with it?
- a. Ford
 - b. Toyota
 - c. Chrysler
 - d. Chevrolet
 - e. Ford

Create a set for the top 10 by average annual premium

Edit Set [Top 10 Premium] X

Name: Top 10 Premium

General Condition Top

☐ None

☒ By field:

Top 10 by

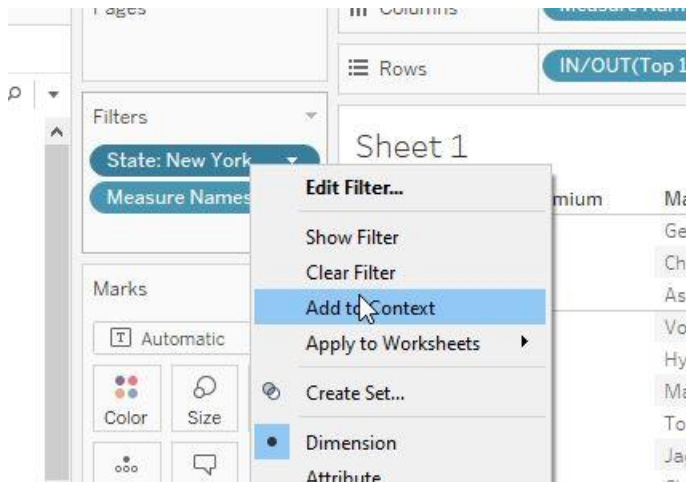
Annual Premium Average

☐ By formula:

Top 10 by

Reset OK Cancel Apply

Filter on State = New York and add to context so that the filter happens *BEFORE* the set. This will give us the top 10 by average annual premium in the state of New York.



Add a count of Unique Policy IDs:

unique policies

```
countd([Policy ID])
```

Final view shows Chevrolet has the most unique policies in the top 10 annual premium set for New York.

Pages

Filters

State: New York

Measure Names

Marks

Automatic

Color

Size

Text

Detail

Tooltip

Measure Values

AVG(Annual Premiu..

AGG(unique policies)

Columns

Measure Names

Rows

IN/OUT(Top 10 Pr..

Make

Sheet 1

In / Out of Top 10 Premium	Make	Avg. Annual Premium	unique policies
In	Geo	747.8	3.0
	Volvo	550.0	3.0
	Hyundai	470.0	3.0
	Maserati	465.8	10.0
	Toyota	448.4	18.0
	Jaguar	442.2	9.0
	Chevrolet	432.0	24.0
	Plymouth	429.0	3.0
	Chrysler	421.5	9.0
	Lexus	417.1	6.0
Out	GMC	402.4	18.0
	Cadillac	390.7	3.0
	Mazda	380.2	8.0
	Subaru	379.4	9.0
	Pontiac	371.7	20.0
	Ford	366.5	48.0
	Volkswagen	363.2	27.0
	Scion	361.8	6.0

27. 1pts Knowledge Groups – A group in Tableau is most similar to which of the following?

- a. Parameter
- b. Set**
- c. Hierarchy
- d. Combined field

A group is similar to a set in that both sets and groups operate on a single field, unlike hierarchies, parameters and combined fields.

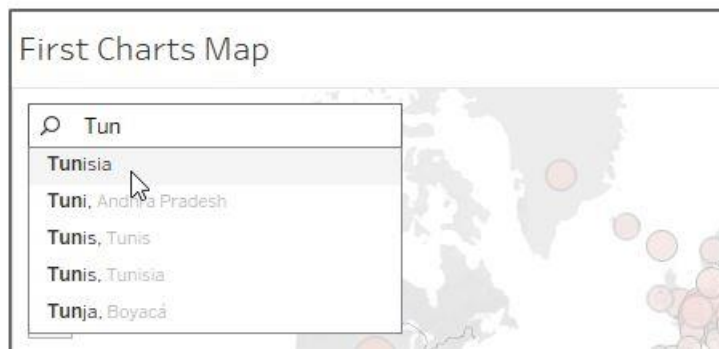
A crucial difference between sets and groups is that with sets, a given value is either IN or OUT of the set, but with groups a value can be in group A, group B, group C, etc. For example, you could create a set called “Mammal” that has fox, bear, cat, and dog IN the set and lizard, flower, tree, beetle OUT of the set. But with Groups you could create a “Mammal” group, a “Reptile” group, an “Insect” group and a “Plant” group.

28. 1pts Knowledge Hierarchies – For which of the following will Tableau automatically create a hierarchy?
- a. Tableau never creates hierarchies automatically – you have to select “create hierarchy” to add them.
 - b. Date fields**
 - c. Geographic fields (Country, State, Zipcode)
 - d. Fields in both the secondary table when your worksheet using a blend
 - e. Fields that are in the right table when your data sources uses a SQL join

Dashboards

29. [4 pt combine worksheets to create dashboard]. Open the CO2 dashboard Which year had the highest total CO2 output for the Tunisia and the two countries that it shares a border with?
- a. 1979
 - b. 2010**
 - c. 2011
 - d. 2003
 - e. None of the above

Search for Tunisia:



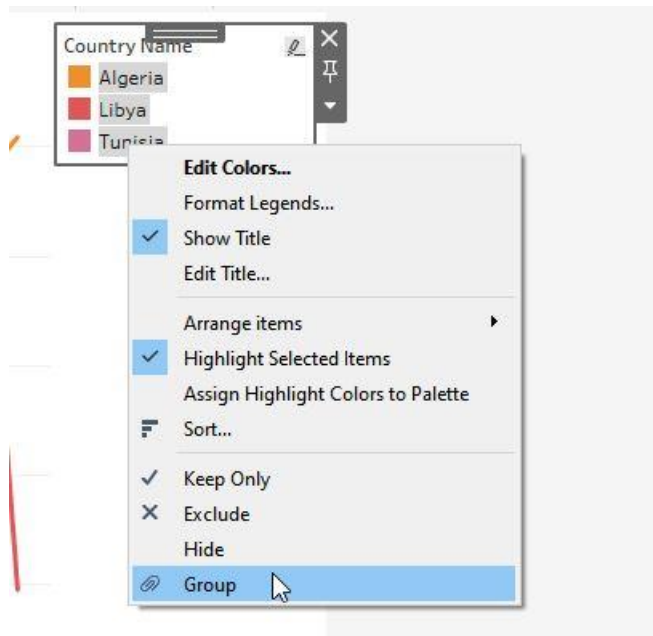
Select the “use as filter” option for the map on the dashboard so that you can filter on Tunisia and countries that border it:



Select the Tunisia, Algeria, and Libya.



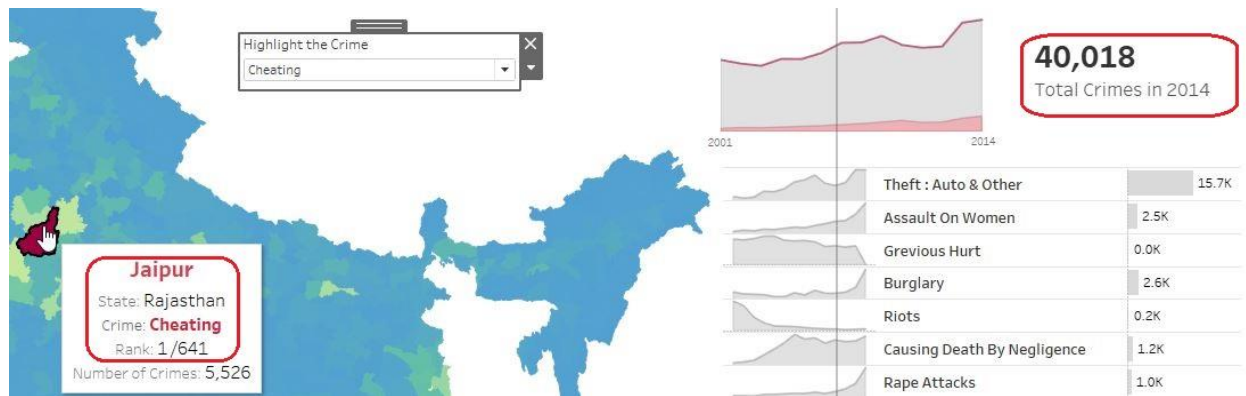
Group the three countries so that they are all shown together in the view:



It should now be clear from the line graph that the peak is in 2010:



30. [3 pt use dashboard] Open the [India Crimes Dashboard](#). Which providence ranks #1 for cheating? How many crimes were committed in that providence overall?
- Bangalore - 49,036
 - Jaipur – 40,108**
 - Mumbai – 59,459
 - Thane – 38,427
 - NCT of Delhi – 50,245



31. [2 pt] You would like to create a dashboard where the user can select one or more values from the header on a bar chart, then have an action that calculates the percent of total represented by the selected values. Which of the following actions can most easily do this?
- Filter Actions
 - Set Actions
 - Change Parameter
 - Change Set Values**
 - Go to URL

See example #4 from this post <http://www.lindseypoulter.com/2018/12/18/getsetgo/?fsi=PHkNzRdf>

32. [1 pt] Which of the following options for sharing a Tableau workbook requires the user to install a Tableau application on their computer in order to view?
- Tableau Online
 - Tableau Server
 - Tableau Reader**
 - Tableau Public
 - None of the Above

Tableau Reader is a free application that can be used to open and see workbooks that have been built in Tableau Desktop. <https://kb.tableau.com/articles/howto/sharing-workbooks-without-tableau-desktop>

33. [1 pt] Which of the following describes the relationship between the default dashboard and the device-specific layouts?
- They all must contain the same worksheets
 - The device-specific layouts can contain either more worksheets or fewer worksheets
 - The default dashboard must contain more worksheets than the device specific layouts
 - The device-specific layouts must contain more worksheets than the default dashboard
 - None of the above**

The device designer video explains, “Only sheets that are put onto the default dashboard will be available for the device-specific layouts. You can remove sheets from a layout, but only what is on the default can be used.” https://www.tableau.com/sites/default/files/device_designer_transcript.pdf So, in other words, the device-specific layouts can have the same worksheets as on the default dashboard or they may have a subset of the worksheets from the default dashboard.

Mapping

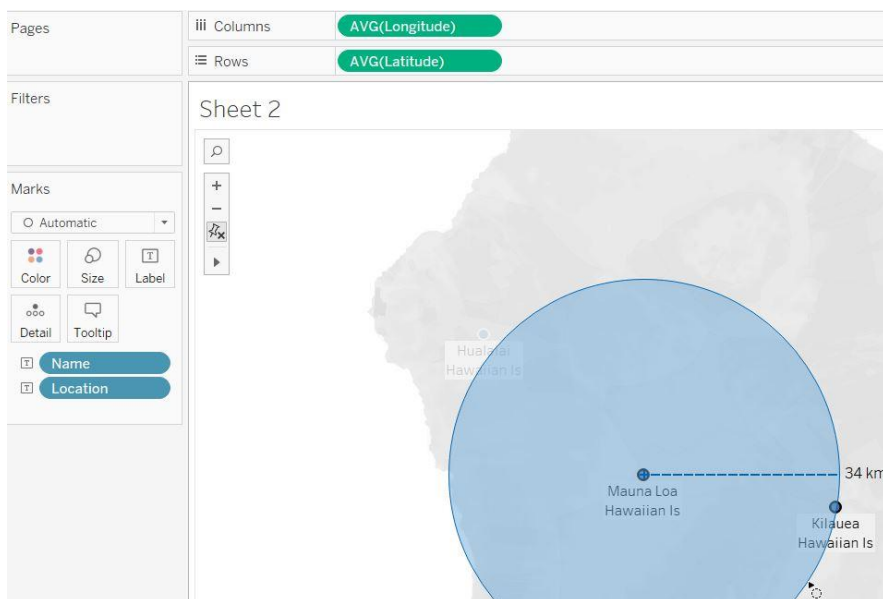
34. 3pts - Radial Selection – distance between two volcanos!

[Radial Selection] Using the significantvolcanoeruptions data set, find the island of Hawaii. Which volcano is closest to Mauna Loa? How many kilometers is it away?

- a. Kilauea, 34 KM
- b. Hualalai, 34 KM
- c. Kilauea, 37 KM
- d. Hualalai, 37 KM

Select Map, Map Options, and set the units to metric.

Add Longitude and Latitude. Do not add the *Longitude (generated)* and *Latitude (generated)* as the generated values are determined by the higher level geographic fields (Country) rather than the exact location of the volcanos. The Latitude and Longitude fields will allow you to locate the volcano exactly.



35. [3pts - Edit Locations in Tableau]

Answer using the “2019 Data” worksheet of the Insurance data Excel file. Correct the city location for the District of Columbia. The city name should be Washington. Which city bordering Washington DC has the closest 2019 annual premium?

- a. Silver Spring
- b. Arlington
- c. Alexandria
- d. Bethesda
- e. Bowie

Use the edit locations options to correct the locations:

Geographic roles

Country/Region: United States

State/Province: State

City: City 3 issues

Match values to locations

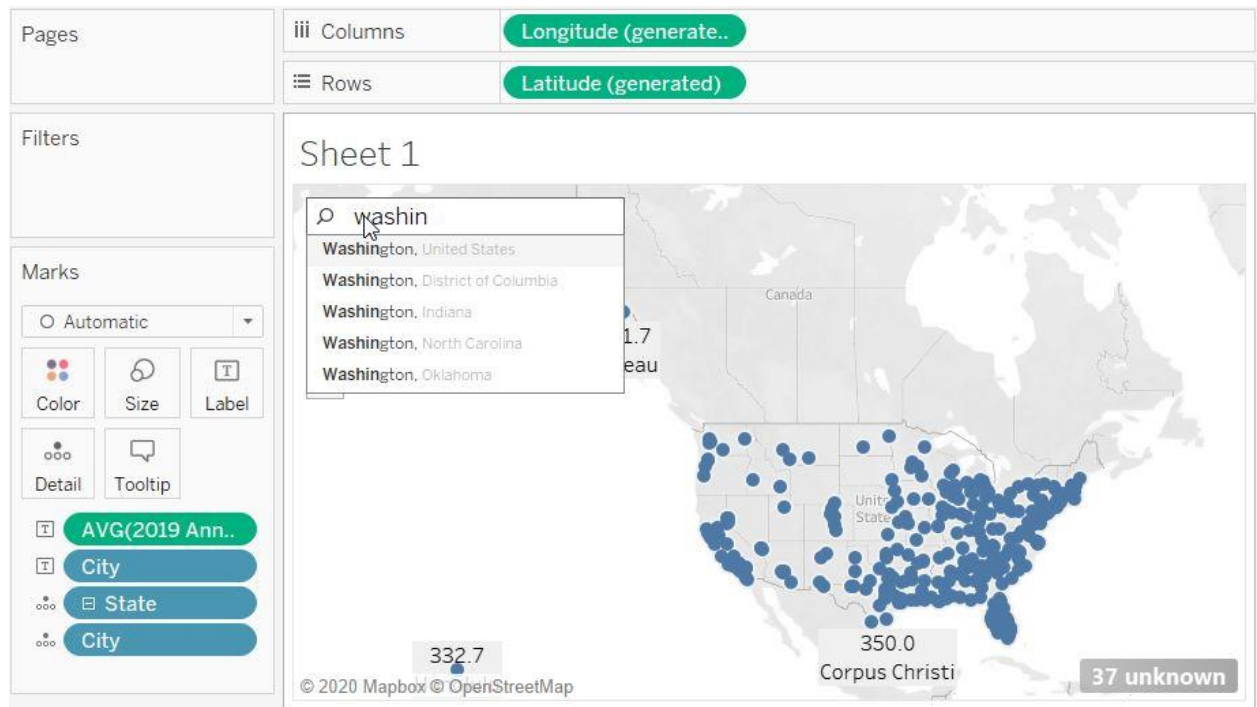
City		State/Province
Your Data	Matching Location	
Merrifield	Unrecognized	
Null	Unrecognized	
Shawnee Mission	Unrecognized	
District of Columbia	Washington	
Abilene	Abilene	
Aiken	Aiken	

☐ Show only unmatched locations in drop down list

Reset Matches OK Cancel

Add city and state as labels. Add average 2019 annual premium.

Search for Washington, District of Columbia



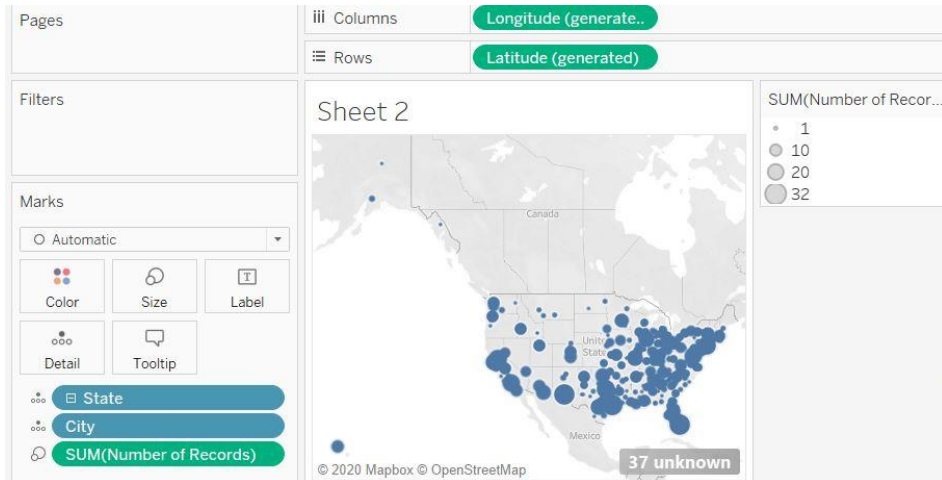
Zoom in:



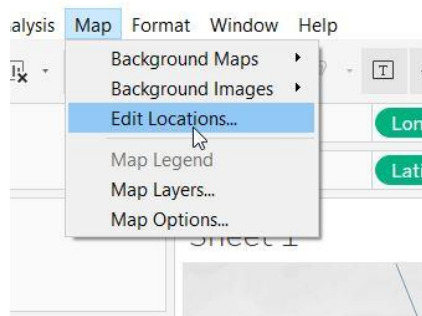
36. [3pts – Map Layers] Answer using the “2019 Data” worksheet from the Car Insurance data excel file. Which city in Orange County, California has the most insurance policies?

- a. Santa Ana
- b. Anaheim
- c. Orange
- d. Newport Beach
- e. Irvine

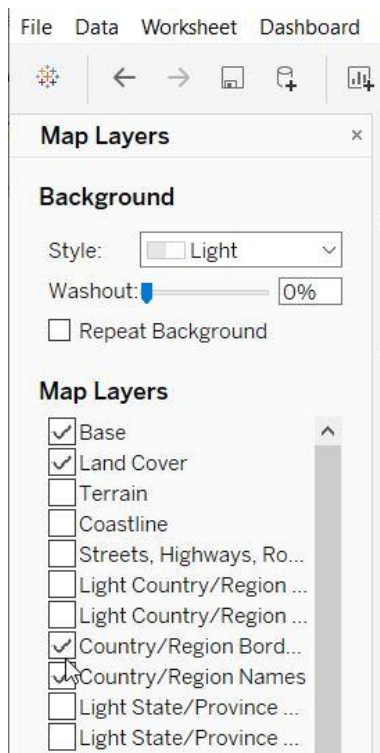
Add city and state to the view. Use Number of Records to count the number of policies:



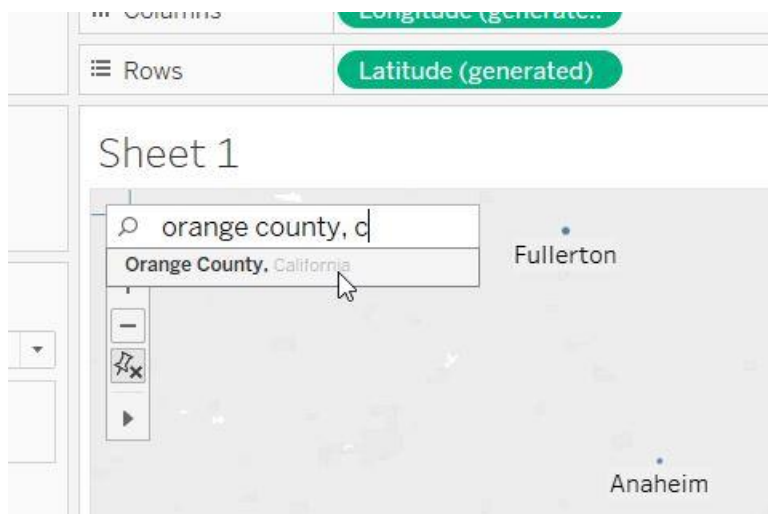
Select Map Layers



Select county names and county borders:



Search for Orange County, California



Irvine has 7 policies, more than any of the other cities in orange county:

