STATISTICAL ANALYSIS IN TABLEAU

LEARNING OBJECTIVES

After this lesson, you will be able to:

1. Explore Tableau's statistical capabilities in the Analytics pane.
2. Build and visualize common statistical models.
3. Develop statistical visualizations to analyze sample data for insights.

DATASETS

In today's lesson, we'll examine how common statistical principles can be applied in Tableau. We will be working with:

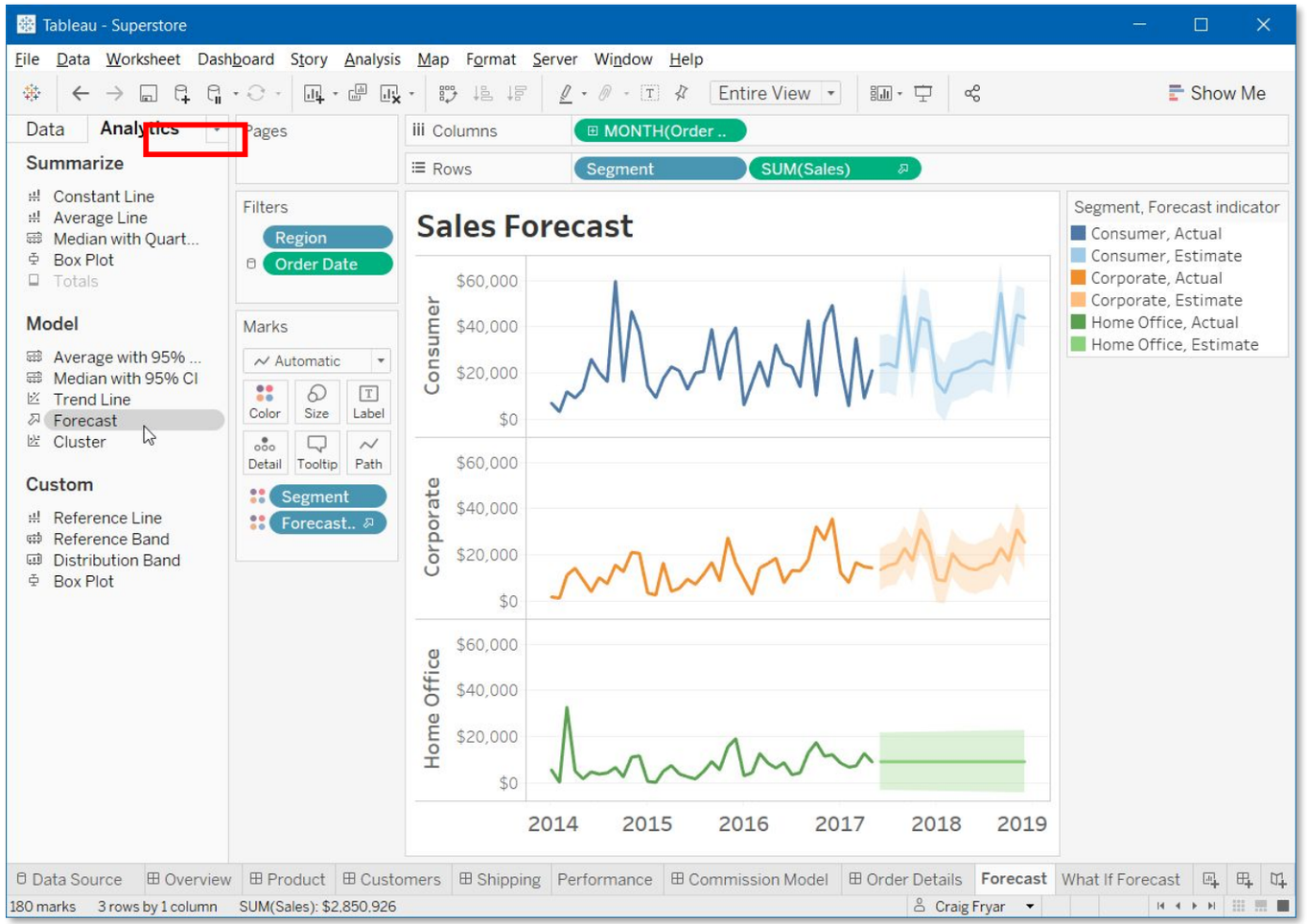
- Two sample data sets, which were included in your Tableau install - Superstore and World Indicators.
- An additional call center data set (excel file to be downloaded).

GUIDED PRACTICE: TABLEAU'S ANALYTICS PANE

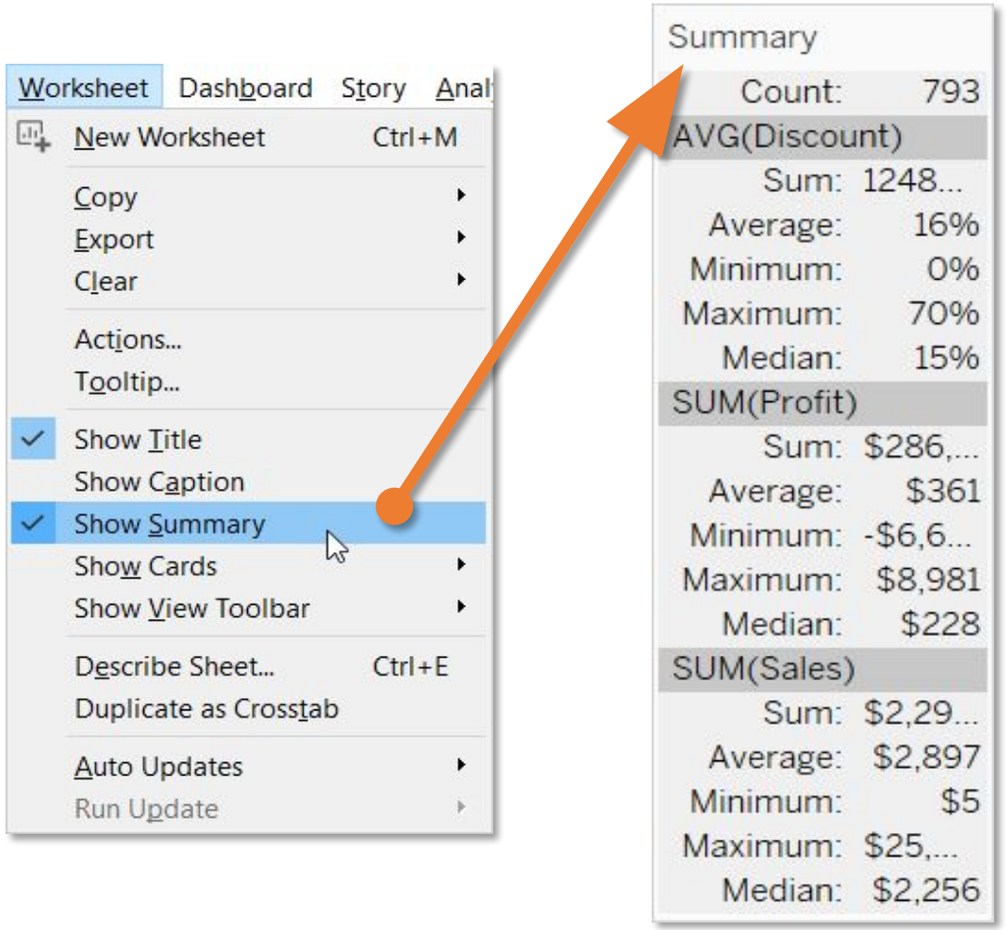
STATISTICAL ANALYSIS IN TABLEAU

Here is an example of Tableau's **Analytics** pane. Some of the statistical functions we'll review in this lesson include:

- Summary statistics.
- Distribution bands.
- Forecasting
- Trend analysis.
- Cluster analysis.
- Box plots.



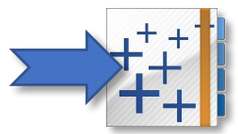
STATISTICAL ANALYSIS IN TABLEAU: SUMMARY STATISTICS



Descriptive statistics are available in all Worksheet menus.

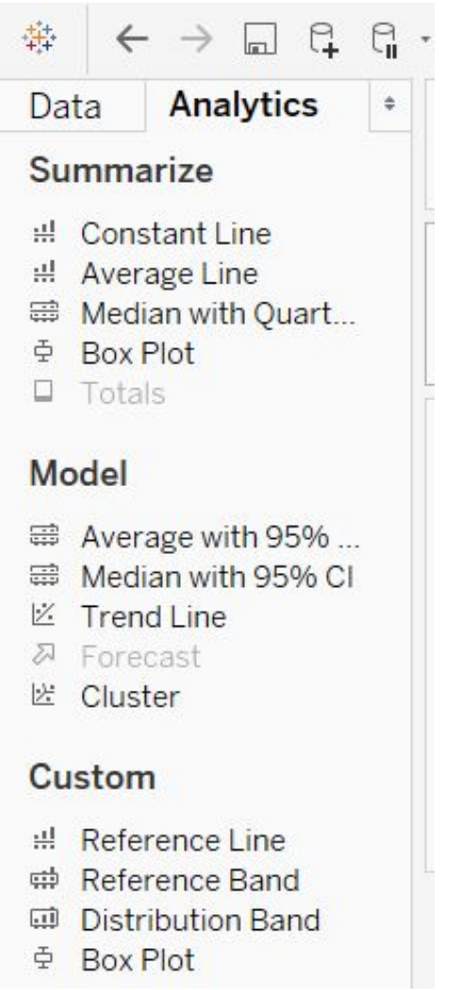
These can display statistics for many different kinds of visualization measures, including:

- Count.
- Sum.
- Average.
- Min.
- Max.
- Median.



TABLEAU'S ANALYTICS PANE

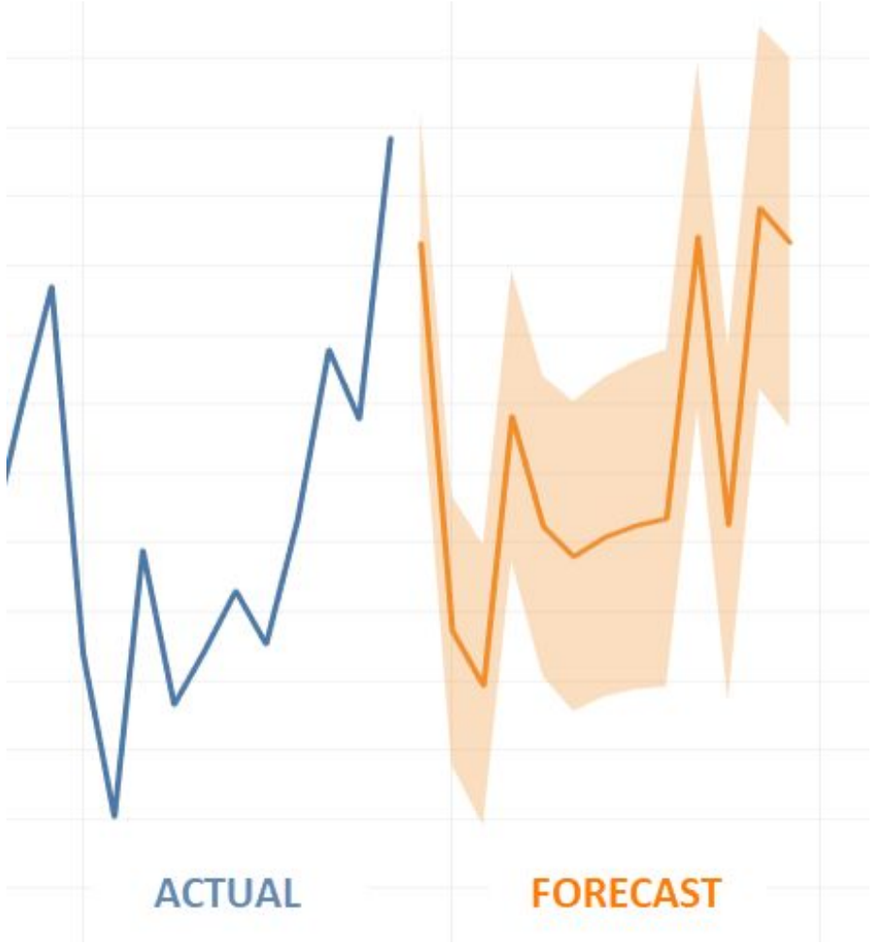
- The **Analytics** pane provides quick-and-easy access to advanced analytical features in Tableau.
- You can drag forecasts, clustering analysis, trend and reference lines, box plots, and more.
- Toggle between the **Data** pane and the **Analytics** pane by clicking one of the tabs at the top of the sidebar.
- There are other ways to add these analyses in Tableau, but the **Analytics** pane allows for drag-and-drop convenience.



STATISTICAL ANALYSIS IN TABLEAU: FORECASTING

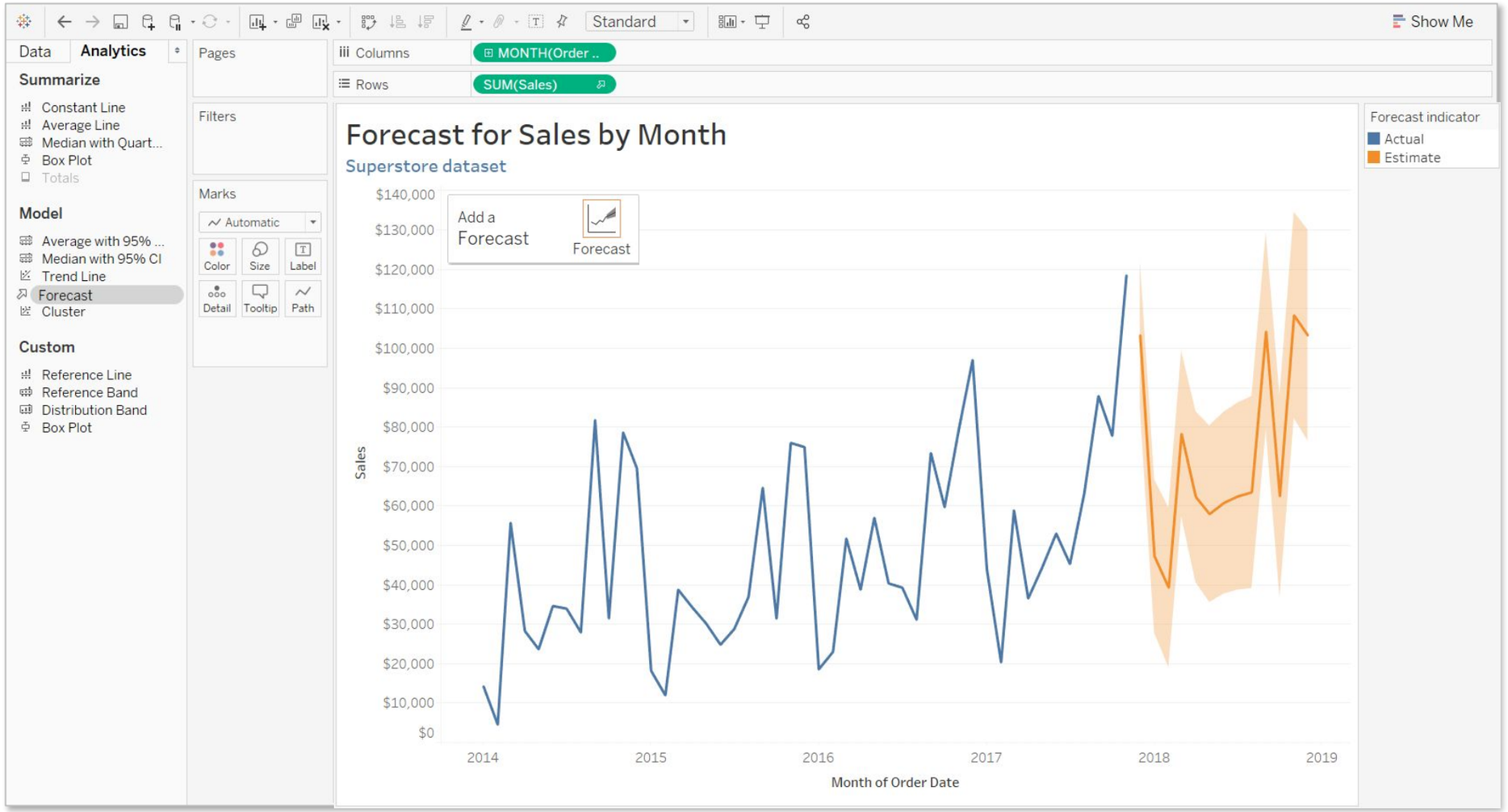
If you're working with quantitative time series data, you can use **forecasting** to create exponential smoothing models.

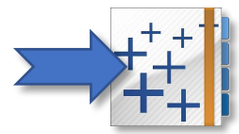
- Future values for a given measure are shown next to the actual values.
- When forecasting, recent data are given slightly more weight.
- Forecasting captures evolving trends and seasonality.



STATISTICAL ANALYSIS IN TABLEAU: FORECASTING

Here's forecasting in action!





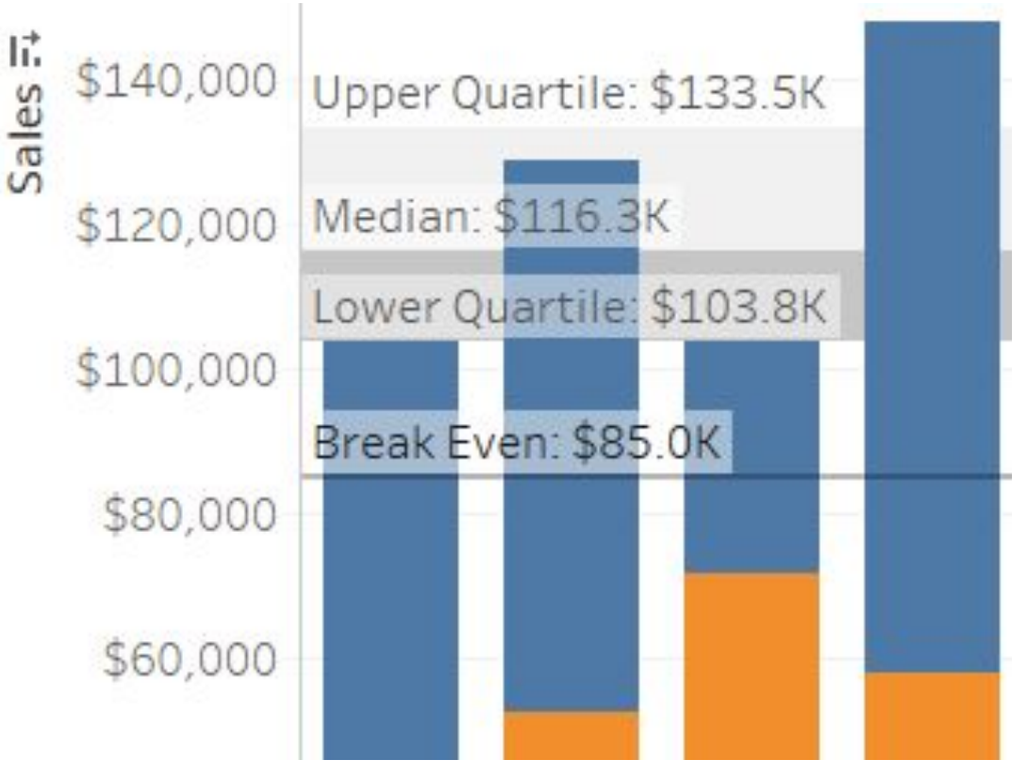
STATISTICAL ANALYSIS IN TABLEAU: REFERENCE LINES AND BANDS

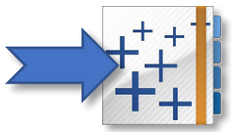
Reference lines can be added to any continuous axis in the view.

- Drag a reference line into the pane to open the edit box.
- Lines may be added by measure, or by table, pane, or cell.

Reference distributions may specify one, two, or more values.

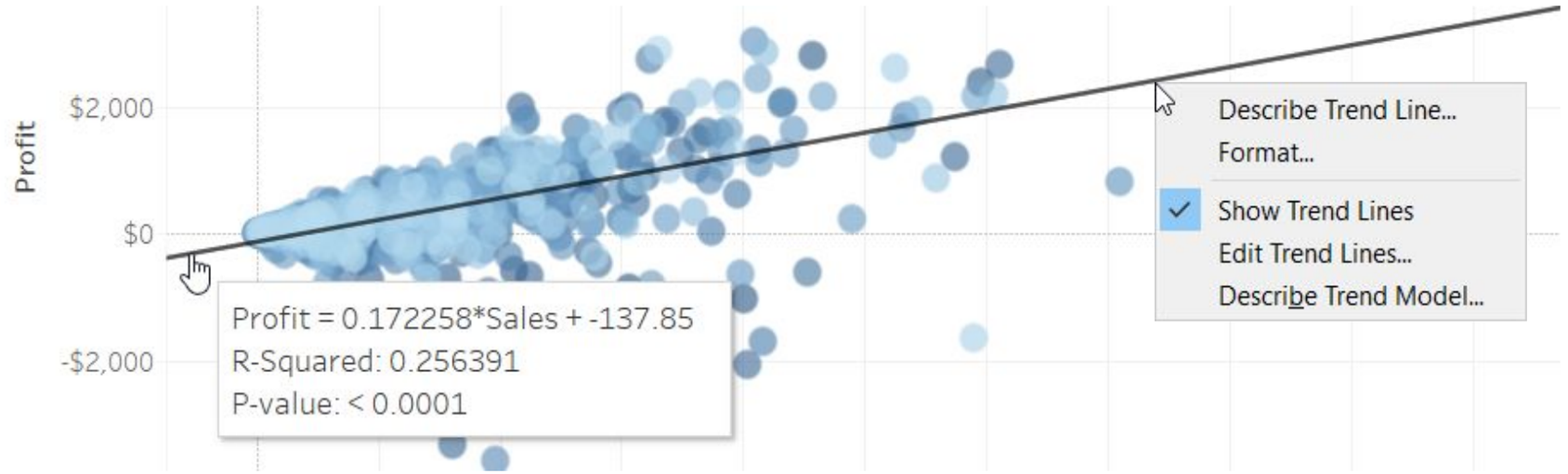
- One value results in a reference line.
- Two or more values create a set of one, two, or more bands.

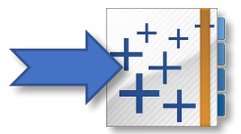




STATISTICAL ANALYSIS IN TABLEAU: TREND LINE ANALYSIS

- You can add one or more **trend lines** in the Tableau view canvas.
- When you add a trend line, the drop options identify your available trend line options: **linear, logarithmic, exponential, and polynomial.***
- Click on a trend line to remove, edit, or to see the model's definition.
- You can remove a trend line by dragging it off the view.



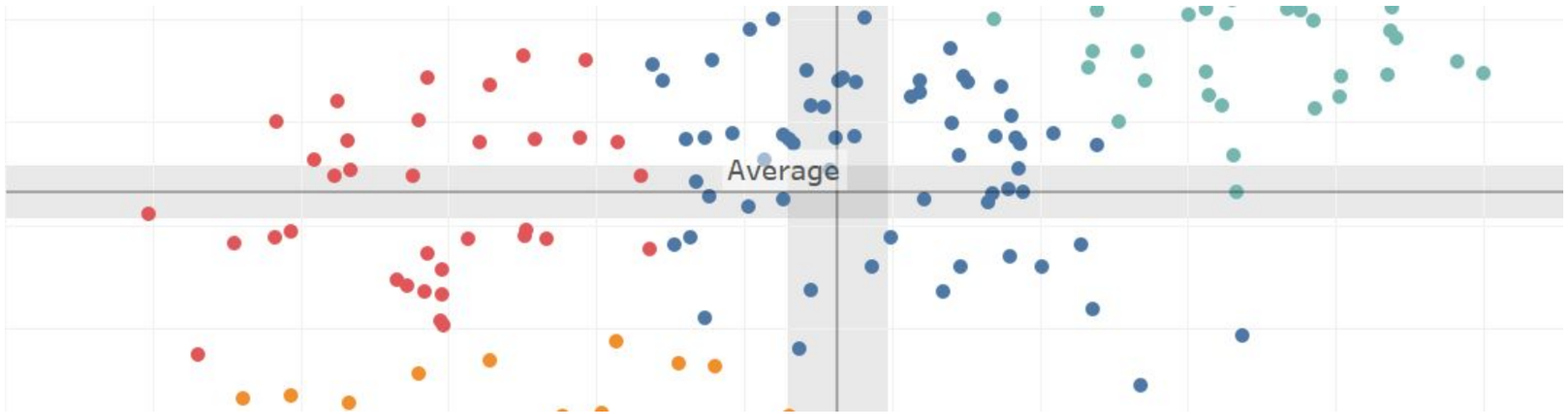


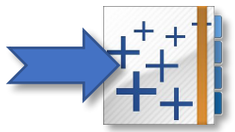
* More information on Tableau trend line models: https://goo.gl/L8bnio.

STATISTICAL ANALYSIS IN TABLEAU: CLUSTER ANALYSIS

Tableau's **clustering** partitions data into statistically similar clusters. This feature can be used to highlight segmentations of marks in a data source.

- Tableau uses the k-means algorithm to group similar marks together into clusters.
- Each cluster has a center called a **centroid**, which is the mean value of all points.
- Clustering is a powerful alternative to manually creating groups or sets.

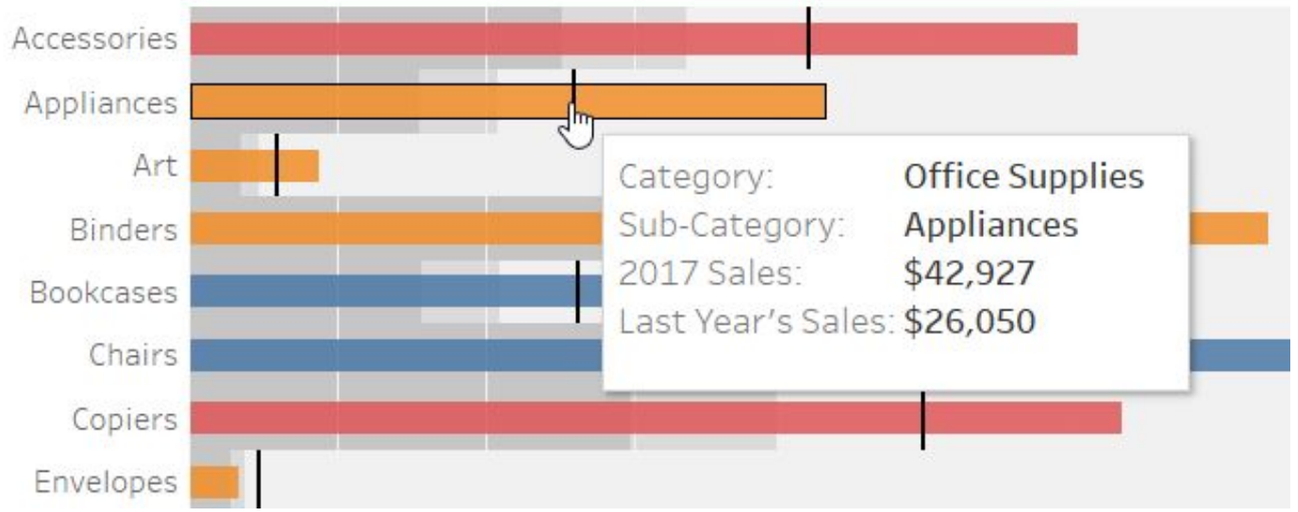


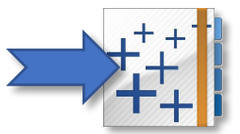


STATISTICAL ANALYSIS IN TABLEAU: BULLET GRAPHS

Reference distributions can also be used to create **bullet graphs**.

- A bullet graph is a variation of a bar graph developed to replace dashboard gauges and meters.
- Bullet graphs are excellent for representing things like targets vs actuals.

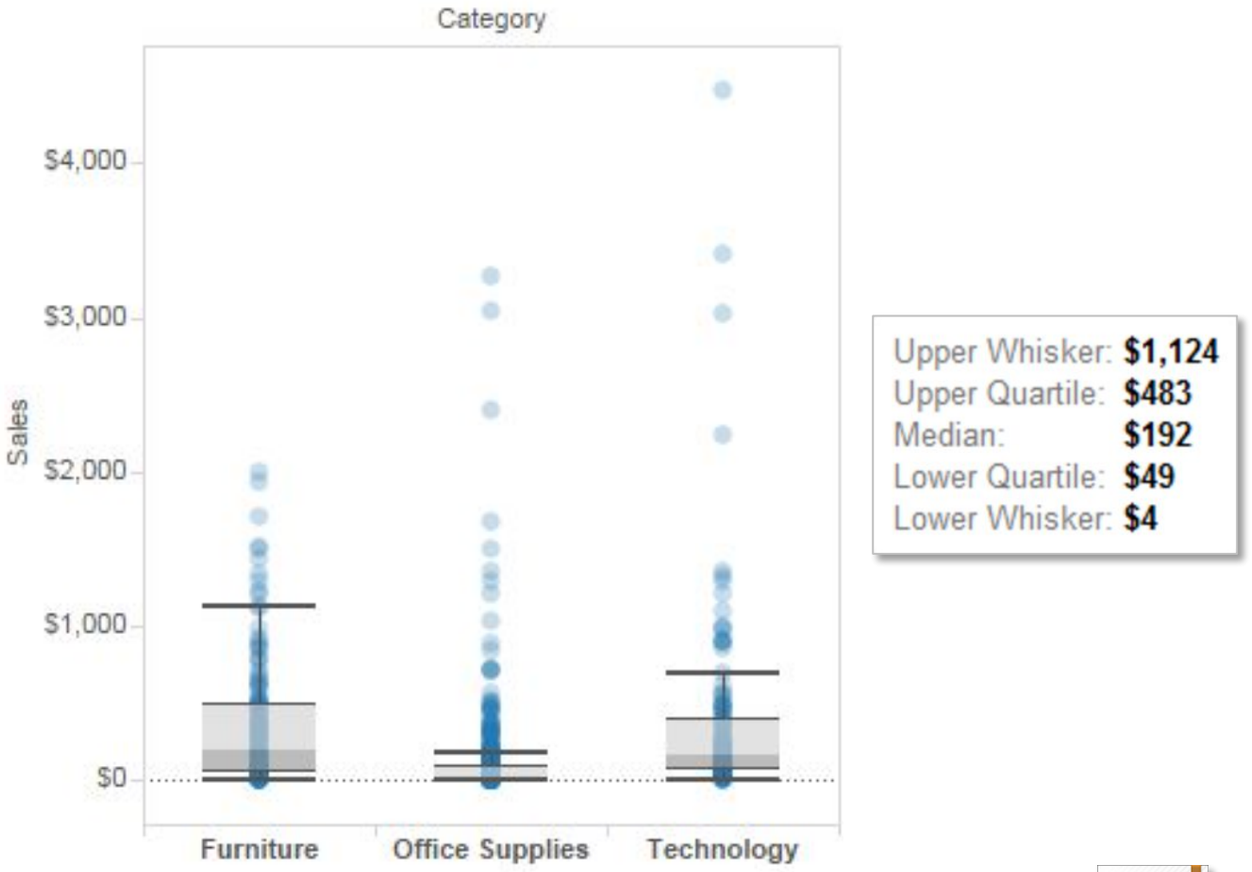


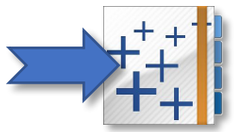


STATISTICAL ANALYSIS IN TABLEAU: BOX PLOTS

Box-and-whisker plots ("**box plots**") create a box around the second and third **quartile** of a range of disaggregated data points.

- Box plots are an effective choice for visualizing distributions.
- Box plots are different from histograms because the "bins" aren't defined as set sizes, but instead by the ***25th***, ***mean***, and ***75th*** *percentiles*.

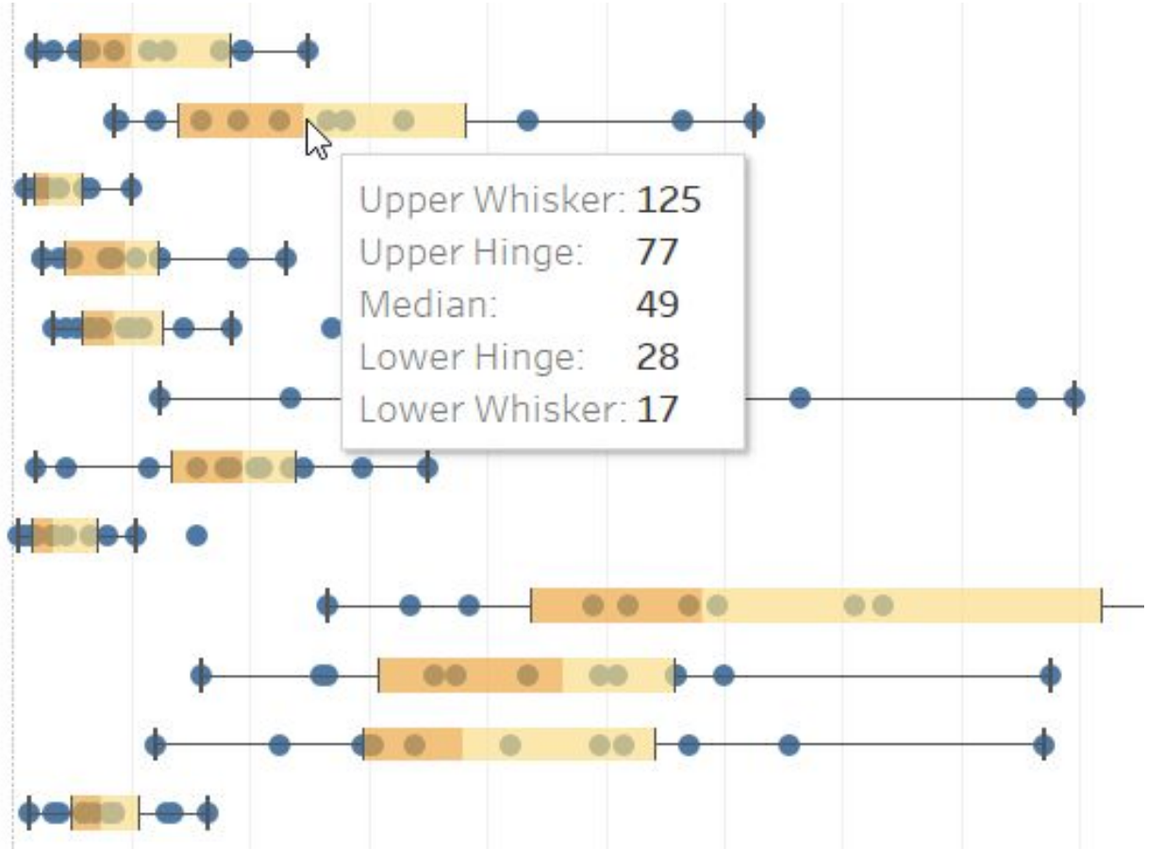


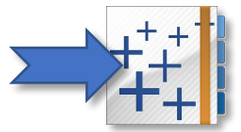


STATISTICAL ANALYSIS IN TABLEAU: BOX PLOTS

Box-and-whisker plots are also great for showing **relative differences** between the distribution of data points for different groups.

- The bottom and top “whiskers” extend to the **first** and **fourth** quartiles.
- Outliers are disregarded in terms of finding the **middle 50** situated around the median.





INDEPENDENT PRACTICE: CALL STATISTICS ANALYSIS

ACTIVITY: STATISTICAL ANALYSIS IN TABLEAU

DIRECTIONS

1. Use the **Statistical_Analysis_in_Tableau_activities_v4** workbook for this activity.
2. Follow the prompts to create the two statistical visualizations located in the independent statistics activity (**blue** tabs).
3. Pair or group up and collaborate with others as appropriate or instructed.

DELIVERABLES

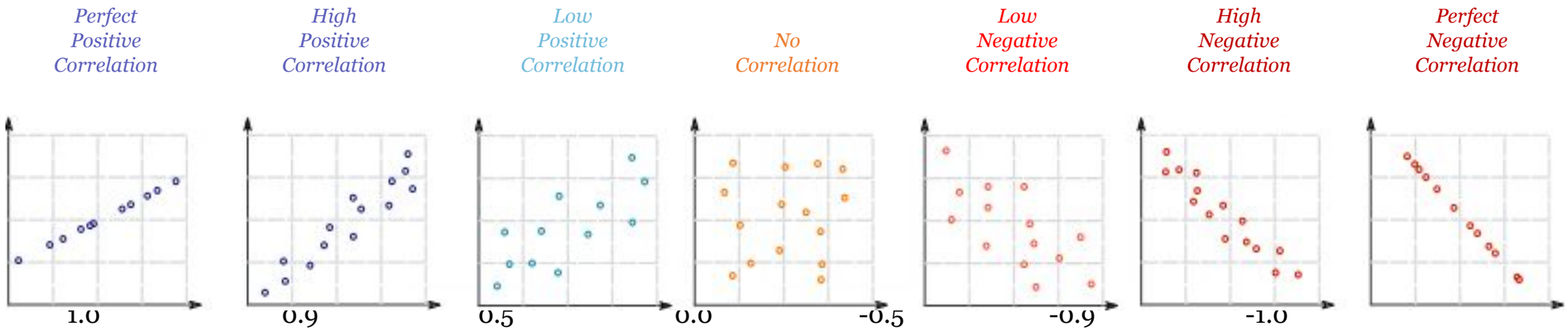
Be prepared to share your efforts and insights.

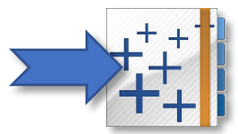
STRETCH EXERCISES: STATISTICS IN TABLEAU

STRETCH EXERCISE 1: SUPERSTORE CORRELATION COEFFICIENT

A **correlation coefficient** quantifies a relationship between variables.

- In a linear correlation, the coefficient quantifies the strength and direction of the correlation between the variables.
- Tableau uses a **Pearson** correlation coefficient, also known as **r**, which measures linear correlation and has a value between ***negative one and one***.

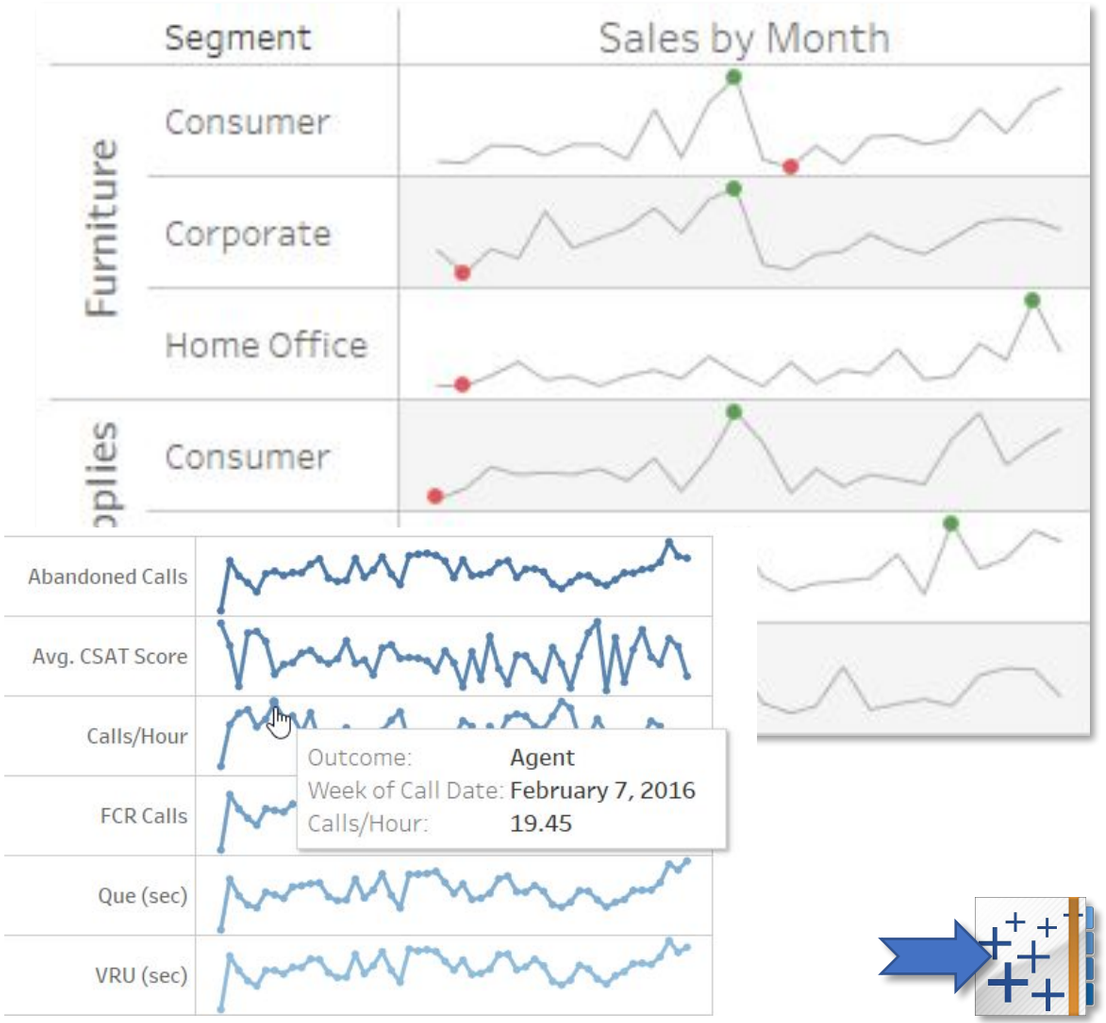




STATISTICAL ANALYSIS IN TABLEAU: SPARKLINES

Edward Tufte introduced **sparklines** in his book "Beautiful Evidence."

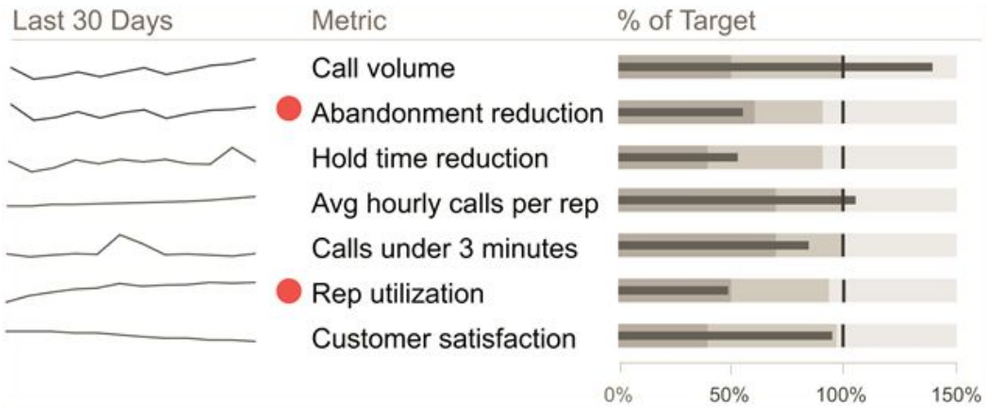
- According to him, a sparkline is a "small intense, simple, word-sized graphic with typographic resolution."
- Sparklines are often incorporated into dashboards and linked to other metrics or statistical elements by **action filters**.



STRETCH EXERCISE 2: CALL CENTER ANALYSIS

1. For our next activity, we'll use the **call_center** data set to statistically explore sample support call KPIs using sparklines.

2. Your goal is to quickly scan for measures of efficiency and effectiveness by using Tableau's statistical analysis tools to inform your answers.



STATISTICAL ANALYSIS IN TABLEAU: STRETCH EXERCISES

DIRECTIONS

Stretch Exercise No. 1

1. Using the **superstore** data, create a correlation coefficient analysis of subcategory sales versus profit.
   a. Segment by region.
   b. Color code the correlations to reflect positive, negative, or no correlation.
   c. Include trend lines.
   d. Filter by region and subcategory.

Stretch Exercise No. 2

1. Using the **call_center** data, develop a set of statistical analyses to measure effectiveness and efficiency of call center agents.
   b. Suggested metrics include customer satisfaction ratings (CSAT), first call issue resolution (FCR), average call times, and overall call volumes.
   c. Develop visualizations and combine them into dashboards and Story Points.

STATISTICAL ANALYSIS IN TABLEAU: STRETCH EXAMPLE SOLUTION

Stretch Exercise No. 3

3. The final tab in this workbook includes a fully featured call center KPI dashboard. It's provided as an *advanced example* of a solution dashboard created with Tableau.

    a. This solution dashboard is intended to be studied, disassembled, and recreated as an independent stretch exercise. **Note**: Intermediate Tableau skills are required.
    b. To understand how this dashboard was created, first click on the **stretch_stats_03** tab, then Unhide All Sheets.
        i. This unhides all of the component visualizations used to build it.
    c. Next, study the components of each visual by selecting Worksheet menu > Describe Sheet.

CONCLUSION

WHAT DID WE LEARN?

To recap, in today's lesson we:

1. Learned how to navigate and use Tableau's Analytics pane.
2. Built and visualized various statistical models in Tableau.
3. Develop statistical visualizations to analyze sample data for insights.

STATISTICAL ANALYSIS IN TABLEAU

Q&A

RESOURCES

- Advanced Analytics With Tableau (Tableau): https://www.tableau.com/learn/whitepapers/advanced-analytics-tableau
- Let's Talk Advanced Analytics (Tableau): https://www.tableau.com/about/blog/2017/4/lets-talk-advanced-analytics-68604
- Using R and Tableau (Tableau): https://www.tableau.com/learn/whitepapers/using-r-and-tableau
- Best-Selling Book "Naked Statistics" (Charles Wheelan): http://a.co/365gXni