quαntcast

Data Visualization Guide

Jun 18, 2014

Katherine Long
klong@quantcast.com

What is Data Visualization?

KEEP IT SIMPLE AND CLEAR

One of the primary goals of visualizing data is to take something complicated or dense and simplify it in order to spot patterns and trends. At a high level, there are two uses for data visualization:

1. exploratory data analysis for yourself.
2. trend or pattern analysis to share with others.

In general, the former drives the latter. However, that doesn't mean the visualizations we build for exploratory analysis are immediately transferable.

To produce content that is accessible to as many people as possible, we should assume our audience:

1. is not familiar with the content and context
2. has little time and limited cognitive bandwidth to process information
3. needs a reason to care

Our visual communications should take these assumptions into consideration during the design phase. Towards those ends, we suggest keeping a few themes in mind when designing data visualizations.

1. minimize design elements
2. simplify content
3. provide context

THINGS TO CONSIDER

Minimize design elements.
In our data visualizations, seek to remove extraneous design elements (within reason) that don't convey relevant information. Design should play a functional and supporting role - to highlight the data and to organize the data. The data should be front and center.

Design elements include not only lines and shapes but text as well. Think of labels, axes labels, annotations, and legends as part of the design.

Simplify content.
Fight the temptation to share all the awesome data that you've found (especially in a single chart). Instead, distill the data down to its most salient point and design your visualization around it. At an immediate glance, the audience should be able to quickly identify the trend or pattern that you're sharing.

If you have a lot of data, consider whether there are natural groupings where you can aggregate without losing meaning or whether you can split the data into multiple charts.

Provide context.
Data in a vacuum has limited usefulness. Consider what relevant comparisons and baselines can help orient the audience to the context and environment your data lives in. For example, how does the content compare over time or against other competitors?

USING TABLES, USING CHARTS

Don't discount tables. Depending on your goal, tables can be a far more effective way to communicate data. In general, a table makes sense if your goal is to show exact values, allow lookup of individual values, or faciliate direct comparisons between two values.

A chart's strength lies in the ability to condense a large amount of data into a format that is easily processed by our brains. Use charts for finding patterns and comparing sets of values.

DECIDING ON A CHART

The type of data that you're visualizing should dictate the type of chart you use. In general, if you have compelling data then don't be afraid of using traditional bar charts. While they're often considered 'boring', they're by far one of the most flexible and effective ways to convey information. Ultimately, if you have boring data, using a fancy chart won't make it any more interesting. It's like putting lipstick on a pig.

In general, the more fancy and/or complicated a chart is, the more you have to depend on an audience to be familiar with the content and context, be patient enough to process the chart, or care deeply about the findings.

Basic Guideline

Color Palette

Use #0096C8 as the primary quantcast color.
Use different colors only when they correspond to differences of meaning in the data.
Do not use the same color to represent different data in the same chart.
Use monochromatic colors for sequential data or data set smaller or equal to 3.
Use heterochromatic colors for categorical data or data set greater than 4.
Reserve grays for neutral data, such as the average or total.

Monochromatic

PRIMARY

0096C8

SECONDARY

00B1D5 A5D7E9

FONT

333333

NEUTRAL DATA

666666 B5B5B5

INTERFACE

F6F6F6

Heterochromatic

PRIMARY

0096C8 00274D 54B948 72CCD2 FDBB30 ED1849

*Color sequence in order of preference

SECONDARY

00B1D5 A5D7E9

8CC97A CCE5C1

9ED9DE D4EDEE

FFCD74 FFE7BE

Fonts

Minimize the number of font size and color variations. Avoid the use of italics unless it's used to emphasize content or to supplement another font in close proximation.

Helvetica
PRIMARY TYPEFACE

Light HELVETICA LIGHT
Regular HELVETICA REGULAR
Bold

Liberation Mono
NUMERICAL TABULAR TYPEFACE

0123456 LIBERATION MONO REGULAR
0123456 LIBERATION MONO BOLD

Arial
DEFAULT ALTERNATIVE TYPEFACE

Regular ARIAL REGULAR
Bold ARIAL BOLD

Graph Selection Matrix

	TIME SERIES	RANK	PART TO WHOLE	DISTRIBUTION	CORRELATION
VERTICAL BAR GRAPH	●	●	●	●	
HORIZONTAL BAR GRAPH		●	●	●	
STACKED BAR GRAPH	●		●		
LINE GRAPH	●			●	
SCATTERPLOT					●
PIE CHART			●		
BOX PLOT				●	

Chart Guideline

Seek to remove extraneous design elements (within reason) that don't convey relevant information. For consistency, adopt the following recommendations:

Chart borders: none
Chart background color: White
Font: Helvetica #333
Gridlines: 2 - 6
Number format: round up numbers to the nearest even number, avoid decimal points for axis, use commas for 4 or more digits
Date format: Month Year (e.g. Sep 2013), Month Date (e.g. Sep 5)
Label: center label to the data point
Time Series Ratio (just graph itself): x-axis > y-axis

FONT

333333

GOLDEN RATIO

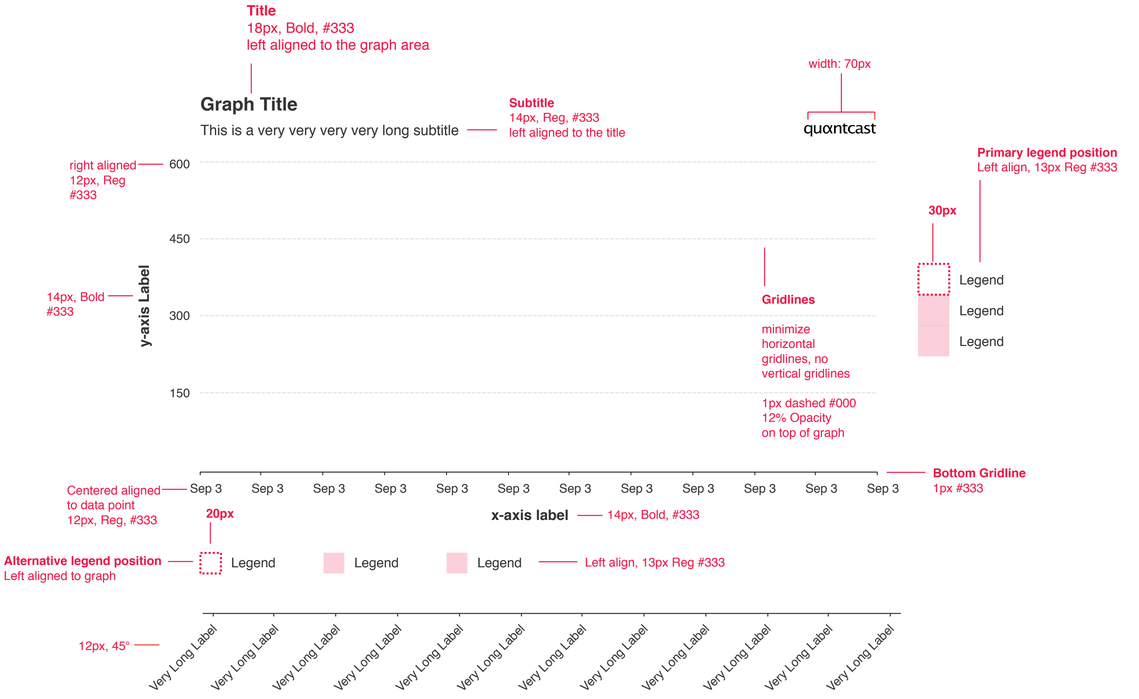
1

1.618

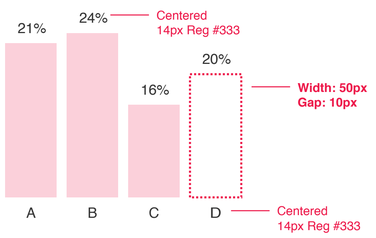
SILVER RATIO

1

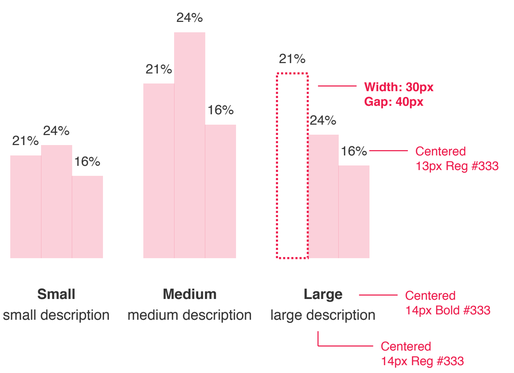
2.414



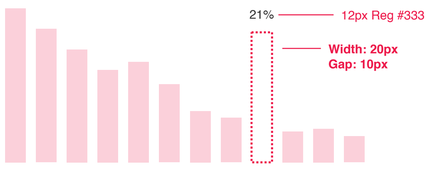
SMALL SET



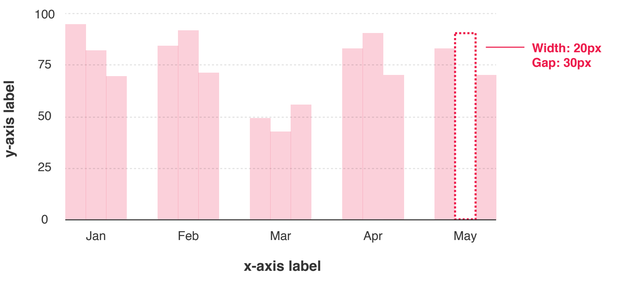
GROUPED SET



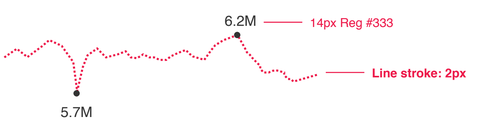
LARGE SET



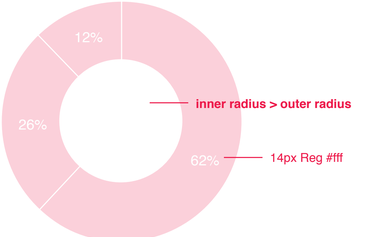
LARGE GROUPED SET



LINE CHART



DONUT CHART



Bar Chart Vertical

Use vertical bar chart for time series, rank, part to whole, and distribution data.
Use horizontal bar chart when the labels are short.
Annotate data in graph if possible. Minimize the use of gridlines, axis labels, and legend when data is annotated in graph.
Limit the number of gridlines between 4 to 6. Please illustrate the 0 axis.

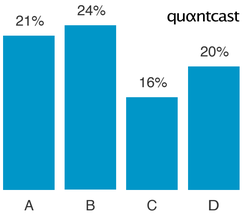
Sequential Data

Monochromatic

*Only use gray to represent average

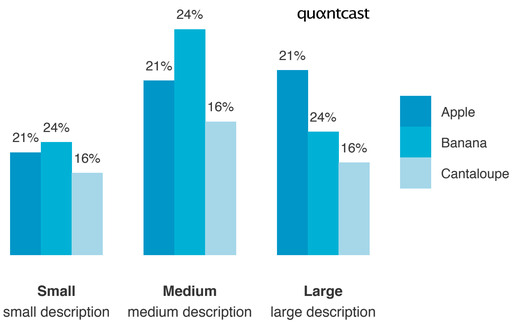
This is an example of a bar chart

This is an example of a bar chart in monochromatic colors



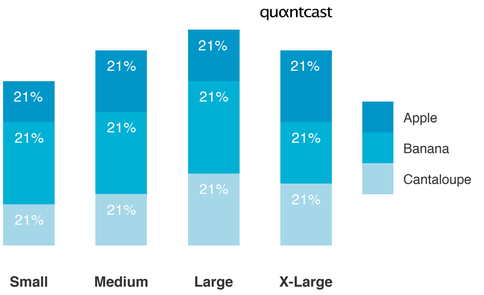
This is an example of grouped a bar chart

This is an example of a grouped bar chart in monochromatic colors



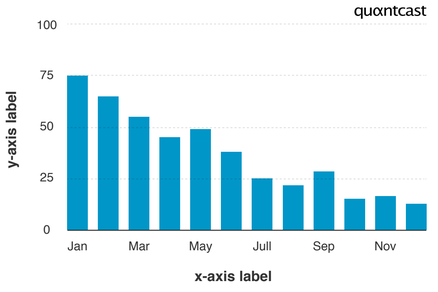
This is an example of a stacked bar chart

This is an example of a stacked bar chart in monochromatic colors



This is an example of a bar chart with gridline

This is an example of a bar chart with gridline in monochromatic colors

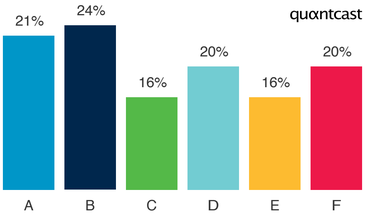


Categorical Data

Heterochromatic

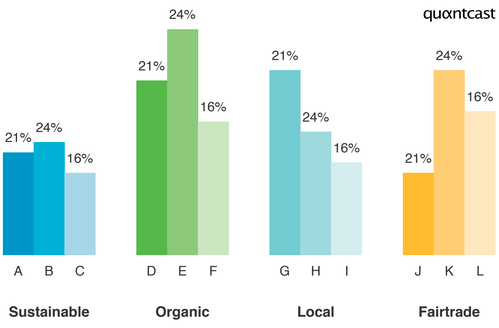
This is an example of a bar chart

This is an example of a bar chart in heterochromatic colors



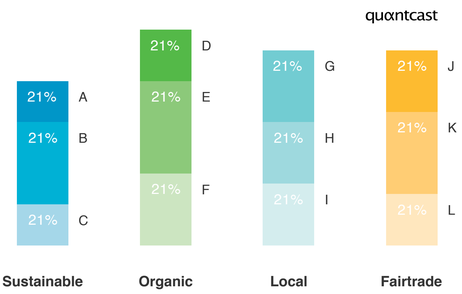
This is an example of grouped a bar chart

This is an example of a grouped bar chart in heterochromatic colors



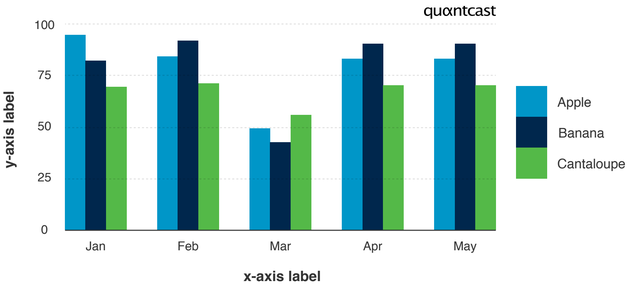
This is an example of a stacked bar chart

This is an example of a stacked bar chart in heterochromatic colors



This is an example of a bar chart with gridline

This is an example of a bar chart with gridline in heterochromatic colors



Bar Chart Horizontal

Use horizontal bar chart for rank, part to whole, and distribution data.
Use horizontal bar chart when the labels are long.
Annotate data in graph if possible. Minimize the use of gridlines, axis labels, and legend when data is annotated in graph.
Limit the number of gridlines between 4 to 6. Please illustrate the 0 axis.

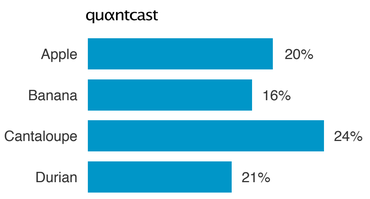
Sequential Data

Monochromatic

*Only use gray to represent average

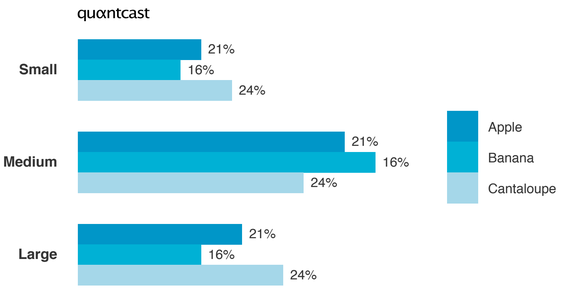
This is an example of a bar chart

This is an example of a bar chart in monochromatic colors



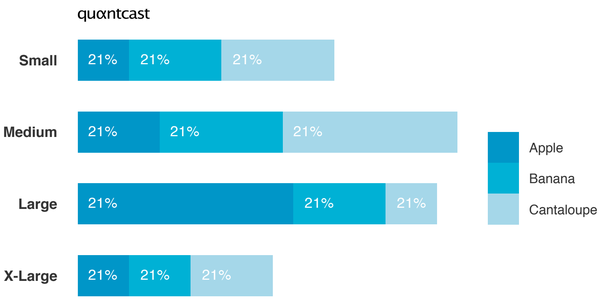
This is an example of a grouped bar chart

This is an example of a grouped bar chart in monochromatic colors



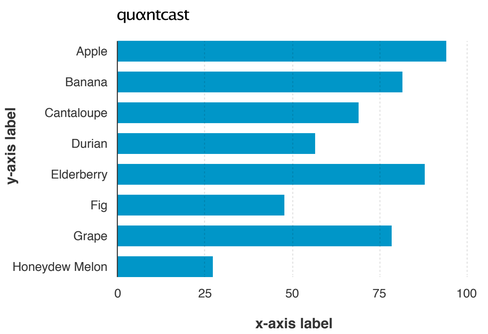
This is an example of a stacked bar chart

This is an example of a stacked bar chart in monochromatic colors



This is an example of a bar chart with gridline

This is an example of a bar chart with gridline in monochromatic colors

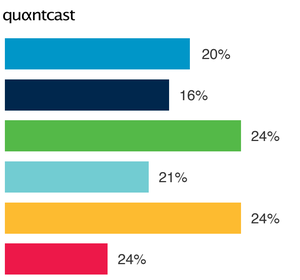


Categorical Data

Heterochromatic

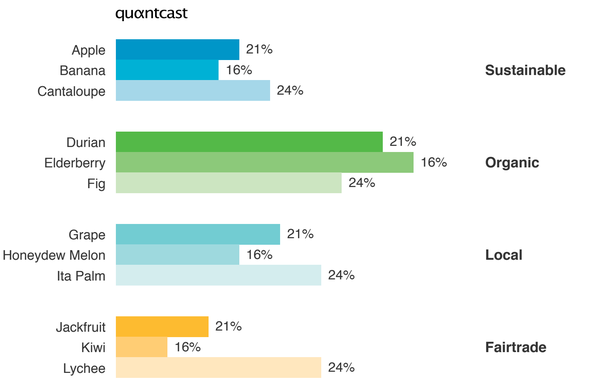
This is an example of a bar chart

This is an example of a bar chart in heterochromatic colors



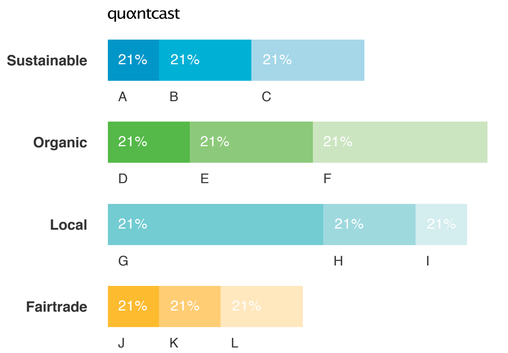
This is an example of a grouped bar chart

This is an example of a grouped bar chart in heterochromatic colors



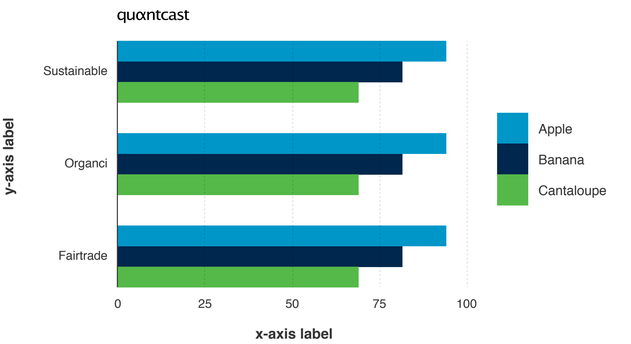
This is an example of a stacked bar chart

This is an example of a stacked bar chart in heterochromatic colors



This is an example of a bar chart with gridline

This is an example of a bar chart with gridline in heterochromatic colors



Line Chart

Use line chart for time series and distribution data.
Annotate data in graph if possible. Minimize the use of gridlines, axis labels, and legend when data is annotated in graph.
Limit the number of gridlines between 4 to 6. Please illustrate the 0 axis.
Stroke width: less than 3 lines: 3px; more than 3 lines: 2px

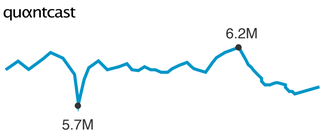
Monochromatic

*Only use gray to represent average

Heterochromatic

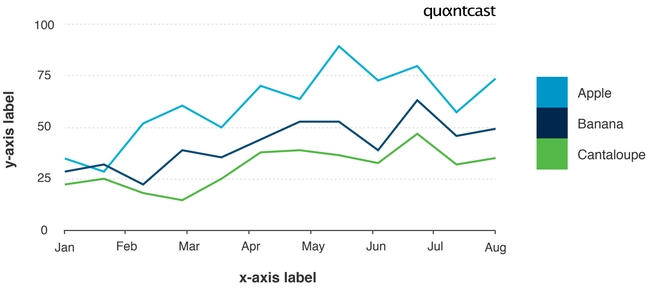
This is an example of a sparkline

This is an example of a sparkline in monochromatic colors



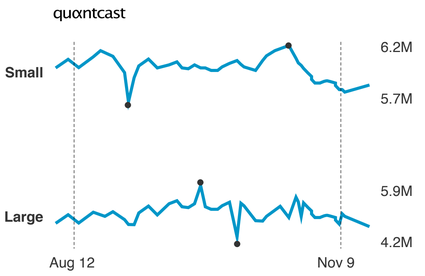
This is an example of a line chart

This is an example of a line chart in heterochromatic colors



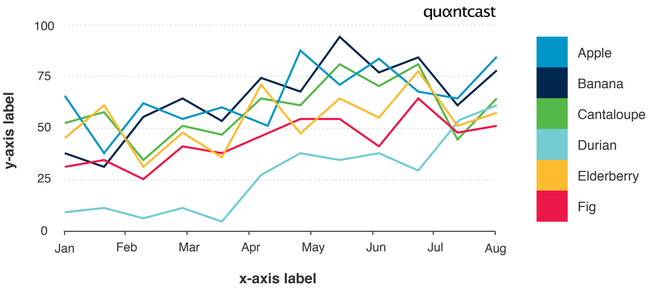
This is an example of a sparkline group

This is an example of a sparkline group in monochromatic colors



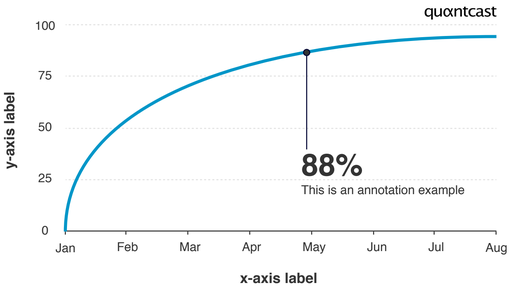
This is an example of a line chart

This is an example of a line chart in heterochromatic colors



This is an example of a simple line chart

This is an example of a line chart in monochromatic colors



Area Chart

Use area chart for time series and distribution data.
Use stacked area chart to highlight the part to whole relationship and emphasize the total. Use the layered area chart for relative comparison between data.
Annotate data in graph if possible. Limit the number of data set to 3.
Limit the number of gridlines between 4 to 6. Please illustrate the 0 axis.

Sequential Data

Monochromatic

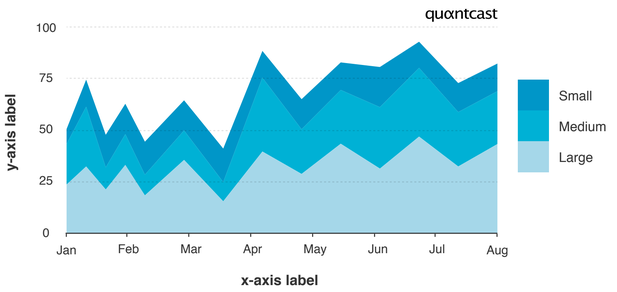
*Only use gray to represent average

Categorical Data

Heterochromatic

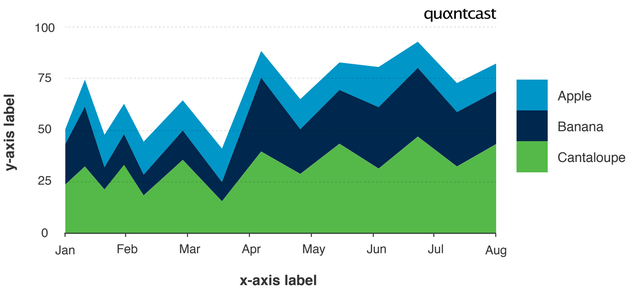
This is an example of a stacked area chart

This is an example of an area chart in monochromatic colors



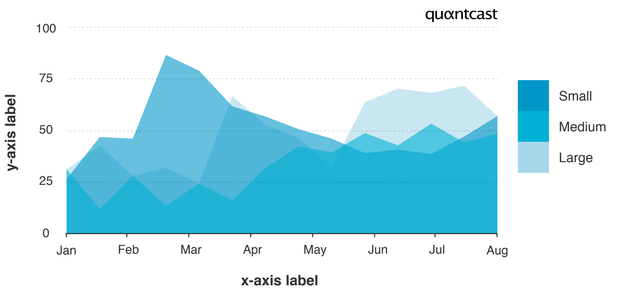
This is an example of a stacked area chart

This is an example of an area chart in heterochromatic colors



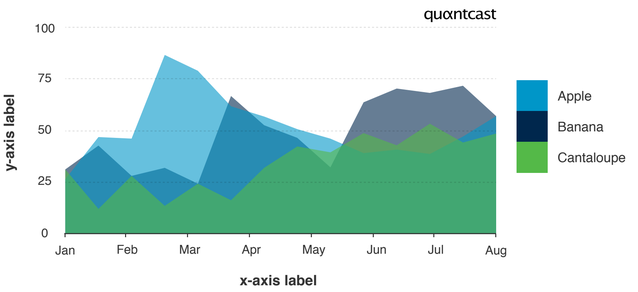
This is an example of a layered area chart

This is an example of an area chart in monochromatic colors at **60% opacity**



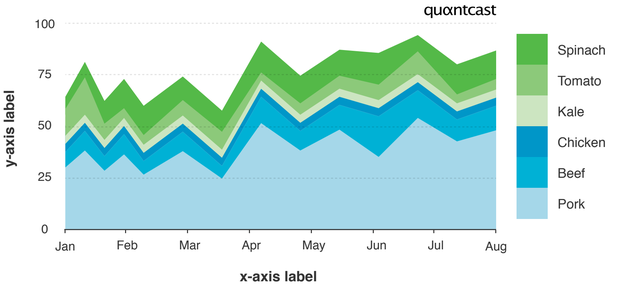
This is an example of a layered area chart

This is an example of an area chart in heterochromatic colors at **60% opacity**



This is an example of a grouped area chart

This is an example of an area chart in grouped monochromatic colors



Segmented Charts

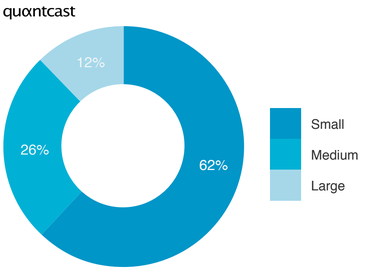
Pie chart and donut chart should be used sparingly since it is hard to gauge and compare each segment's relative size.
We recommend donut charts over pie charts. Donut charts are cleaner and visualize the relative size of each segment more effectively than pie charts.
Minimize the number of segments, we recommend 2-3 segments at most.

Sequential Data

Monochromatic

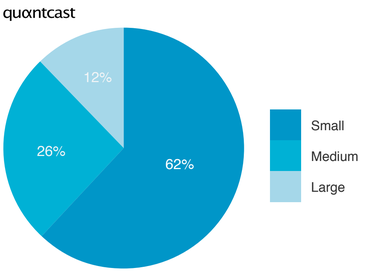
This is an example of a donut chart

This is an example of a donut chart in monochromatic colors



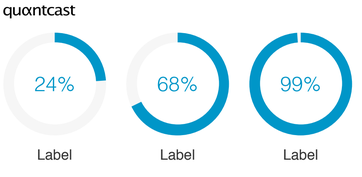
This is an example of a pie chart

This is an example of a pie chart in monochromatic colors



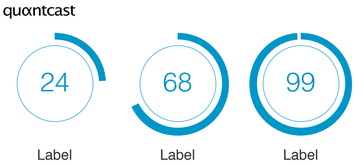
This is an example of a gauge chart

This is an example of a gauge chart in monochromatic colors



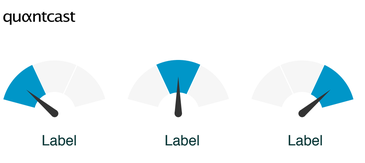
This is an example of a gauge chart

This is an example of a gauge chart in monochromatic colors



This is an example of a gauge chart

This is an example of a gauge chart in monochromatic colors

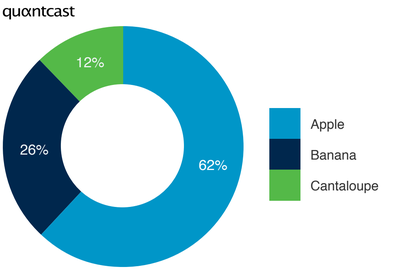


Categorical Data

Heterochromatic

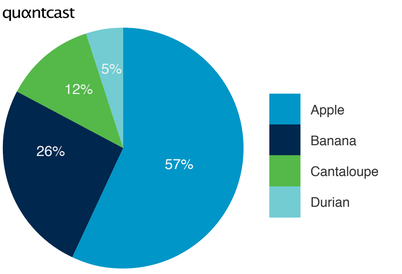
This is an example of a donut chart

This is an example of a donut chart in heterochromatic colors



This is an example of a pie chart

This is an example of a pie chart in heterochromatic colors

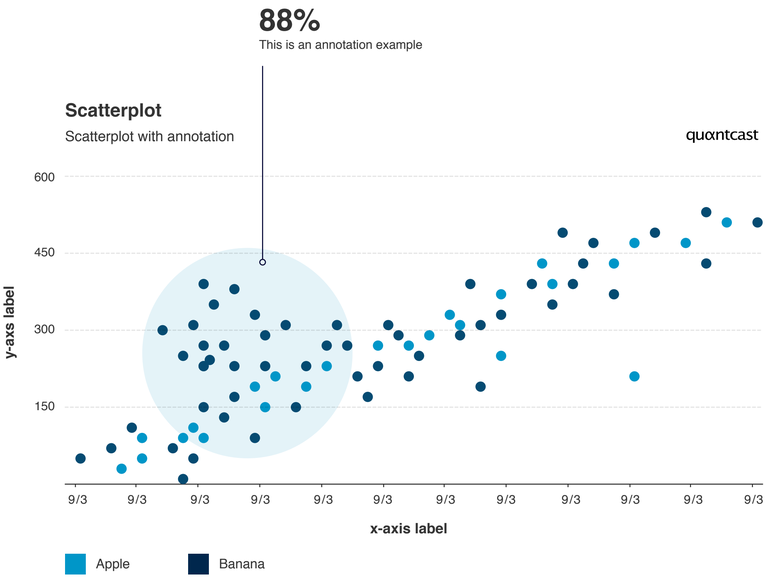


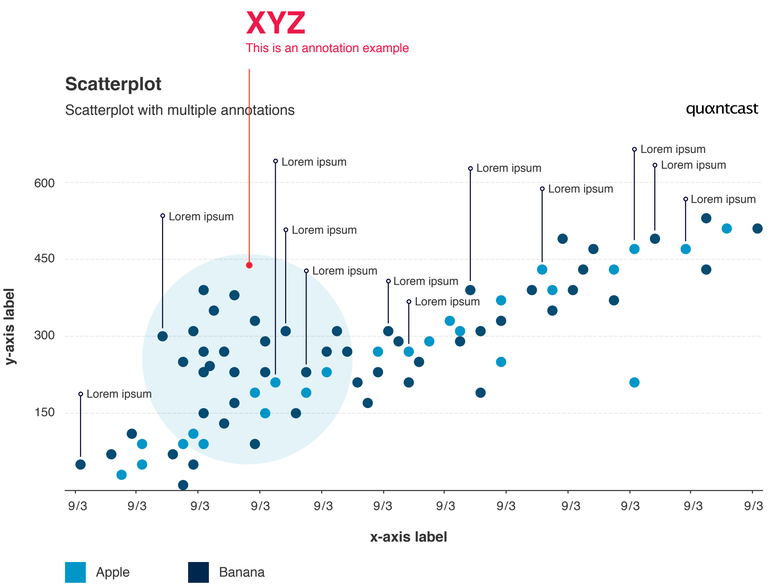
Others

Annotations

Font: Helvetica 12px, regular or bold;
Secondary font: Helvetica 30px, bold
Lines: Vertical or horizontal only
Colors:

333333 ED1849





Tables

Tables are an effective way to show precise values, allow lookup of individual values, or facilitate direct comparisons between two values.

Again, design elements should be minimized so that the data is easily accessible.

- Numerical data / Percentage: right aligned
- Wrap longer table headers to 2 lines max, otherwise truncate
- Always add padding for each cell

Colors:

333333 F6F6F6

Helvetica, 12px, bold, #333

TABLE HEADER	TABLE HEADER	TABLE HEADER	TABLE HEADER	TABLE HEADER
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10

Always add padding for each cell

Text: left align

#f6f6f6

Libration Mono, 13px, #333

Number: right align

border top: #eee

TABLE HEADER	TABLE HEADER	TABLE HEADER	TABLE HEADER	TABLE HEADER
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10
Text	Text	0.10	0.10	0.10

border bottom: #eee