

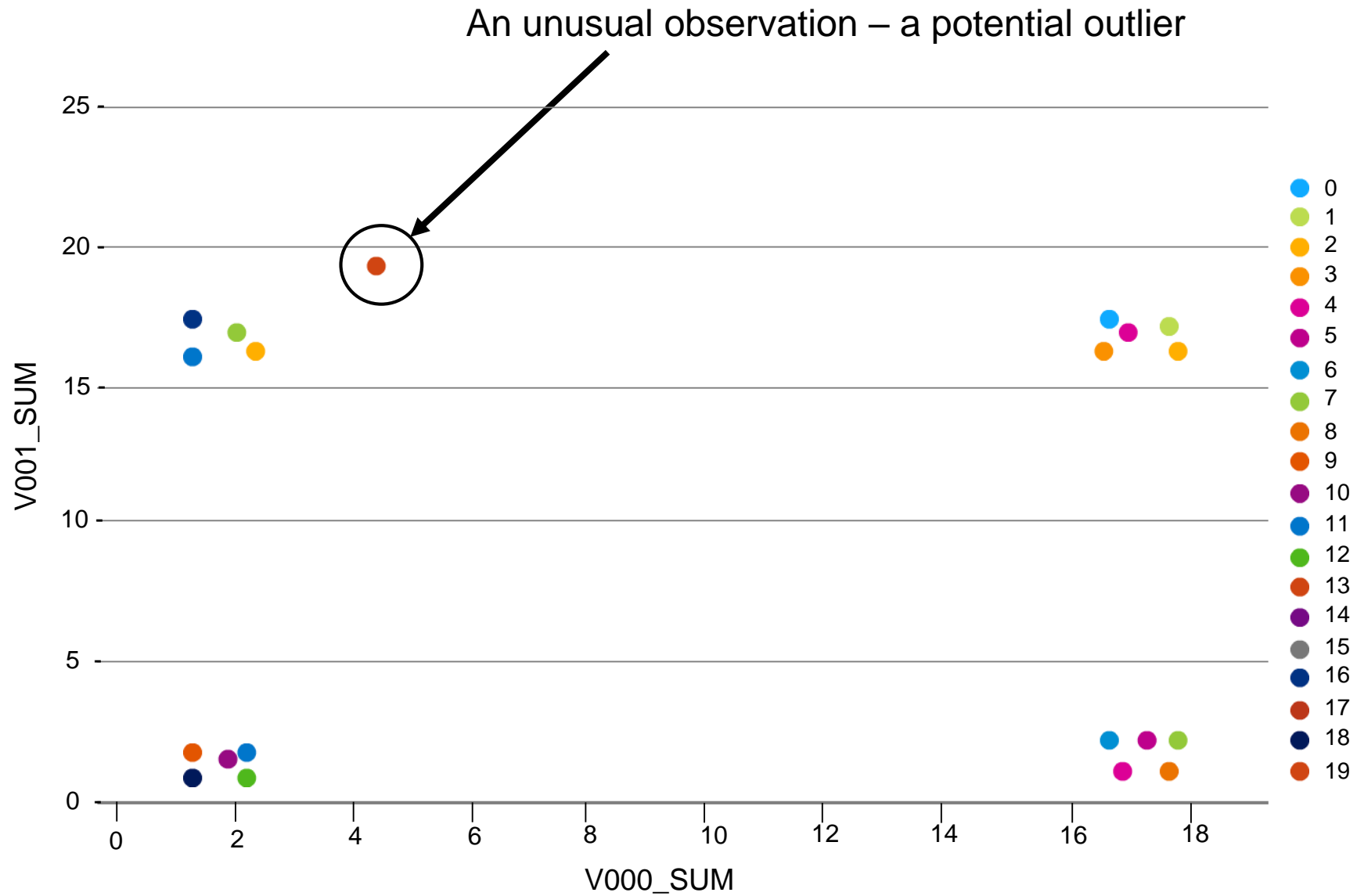


Week 2: Descriptive Statistics

## Unit 6: Outliers

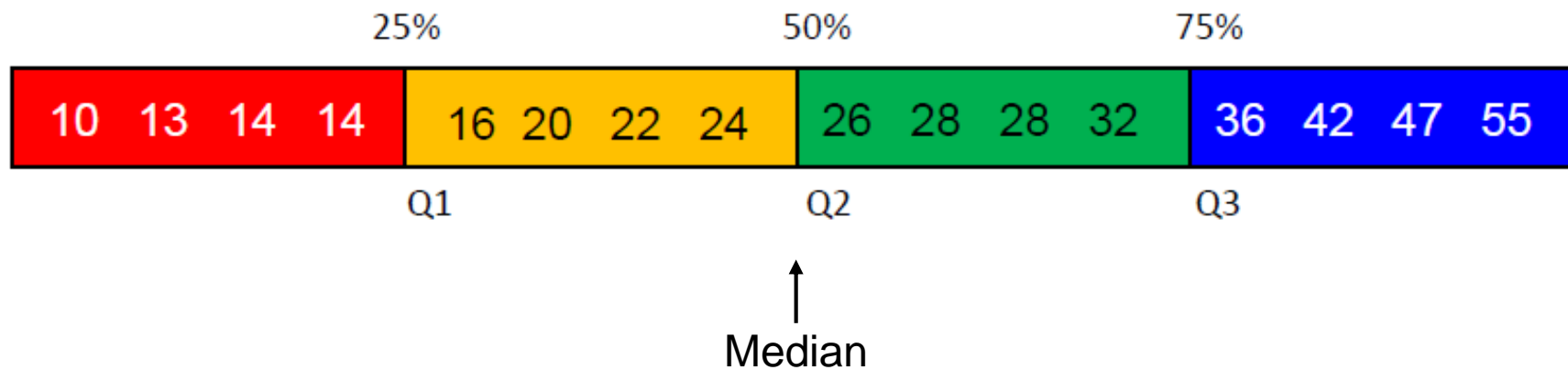
# Outliers

## Introduction

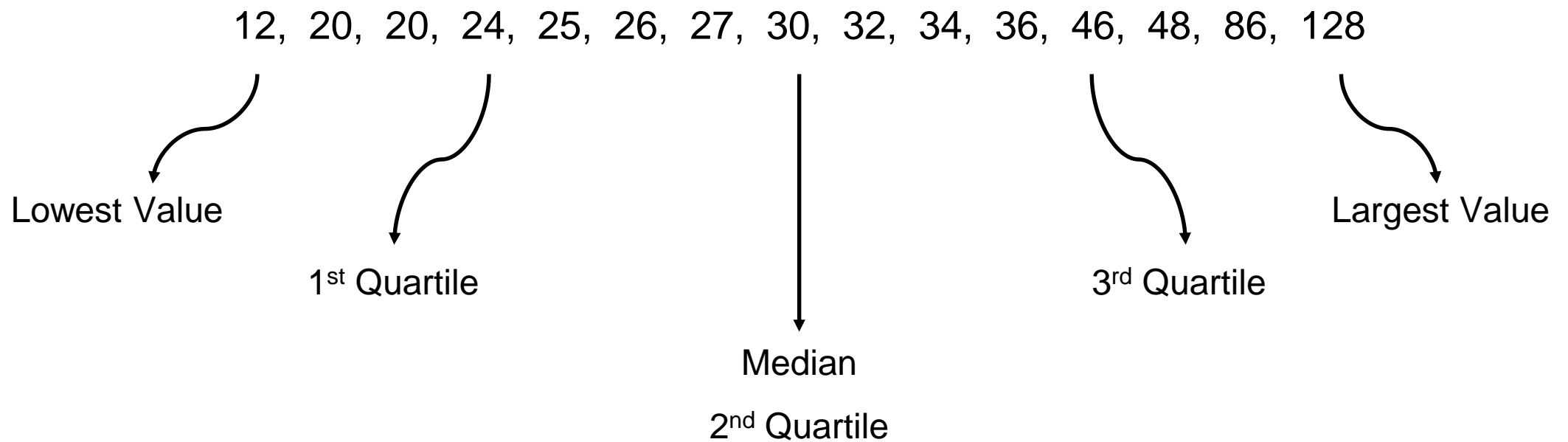


## Understanding quartiles

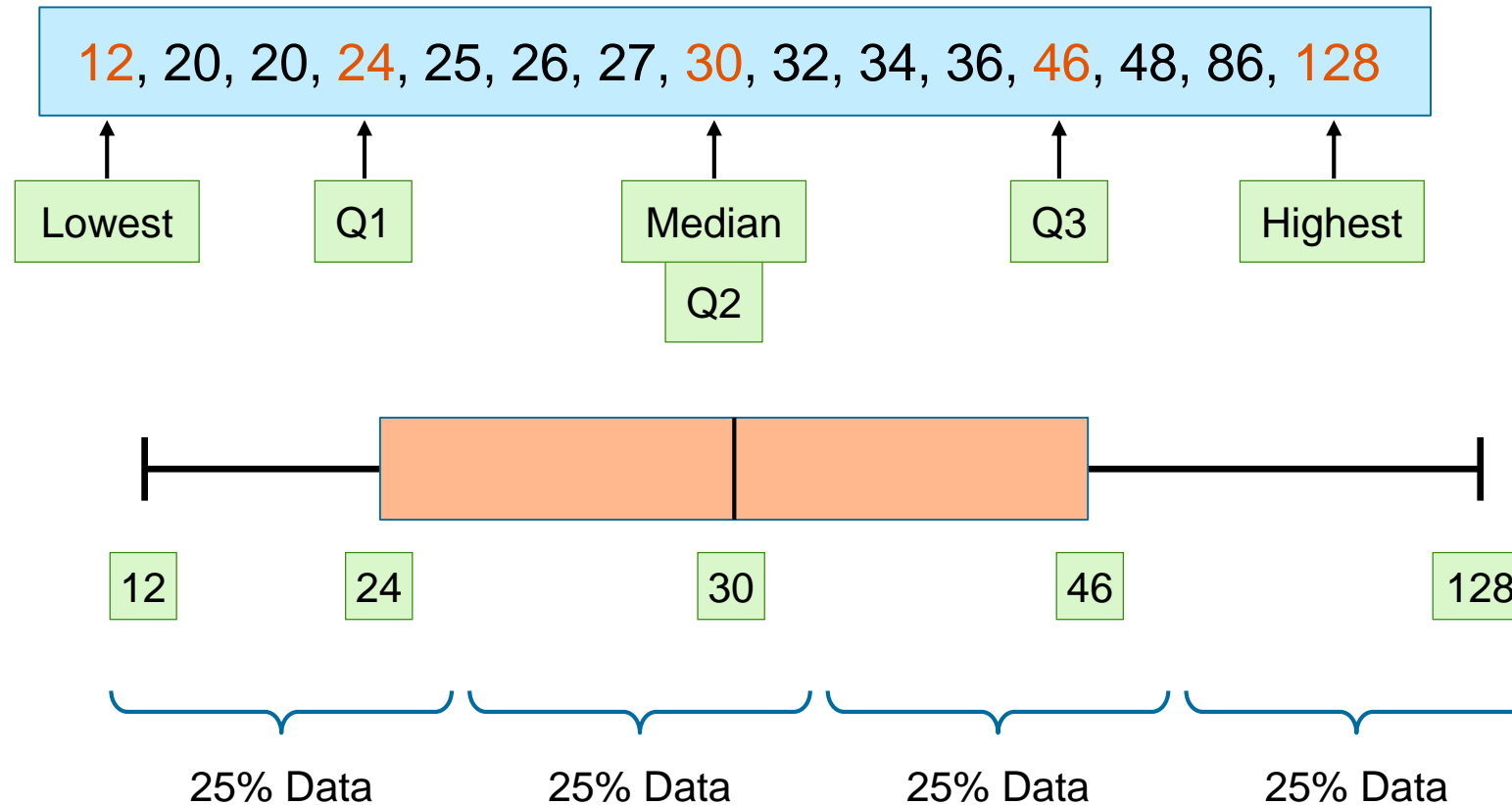
Quartile	Percentile
1 <sup>st</sup> Quartile	25 <sup>th</sup> Percentile
2 <sup>nd</sup> Quartile	50 <sup>th</sup> Percentile
3 <sup>rd</sup> Quartile	75 <sup>th</sup> Percentile



## Quartiles

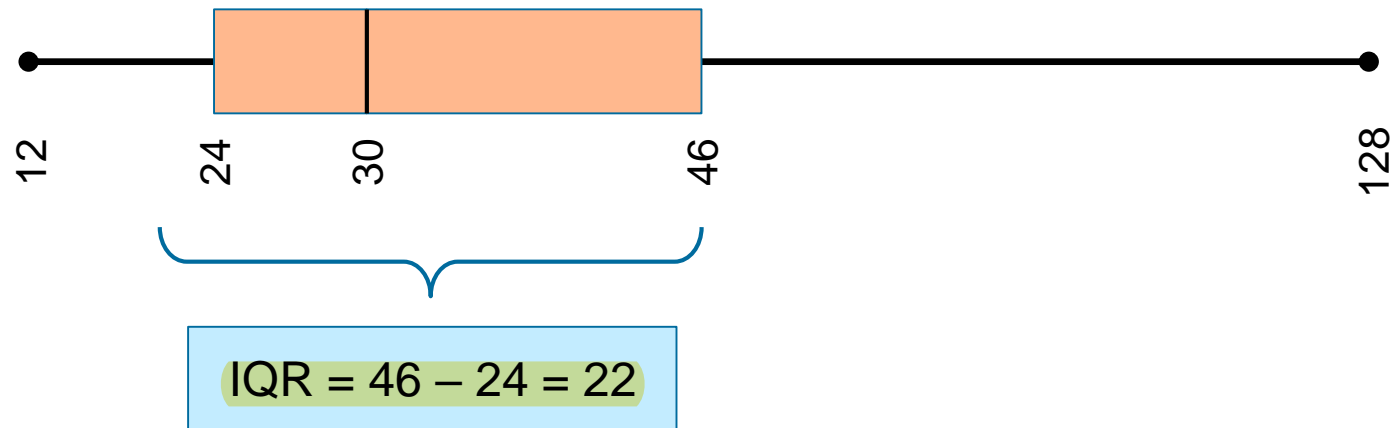


## Detecting outliers using a box plot



## Interquartile range (IQR)

Interquartile Range =  $Q3 - Q1$



## Outliers

### How to identify outlier and extreme value

- Fences are usually found with the following formulas:

$$\text{Upper fence} = Q3 + (1.5 * IQR)$$

$$\text{Lower fence} = Q1 - (1.5 * IQR)$$

- Sometimes you will see reference to “inner fences” and “outer fences”.

$$\text{Lower inner fence: } Q1 - (1.5 * IQR)$$

$$\text{Upper inner fence: } Q3 + (1.5 * IQR)$$

$$\text{Lower outer fence: } Q1 - (3 * IQR)$$

$$\text{Upper outer fence: } Q3 + (3 * IQR)$$

#### Outlier

Left Inner Fence	= $Q1 - 1.5 (IQR)$	→	$24 - 1.5 (22)$	= -9
Right Inner Fence	= $Q3 + 1.5 (IQR)$	→	$46 + 1.5 (22)$	= 79

A value that is less than -9 or greater than 79 is an outlier

#### Outlier

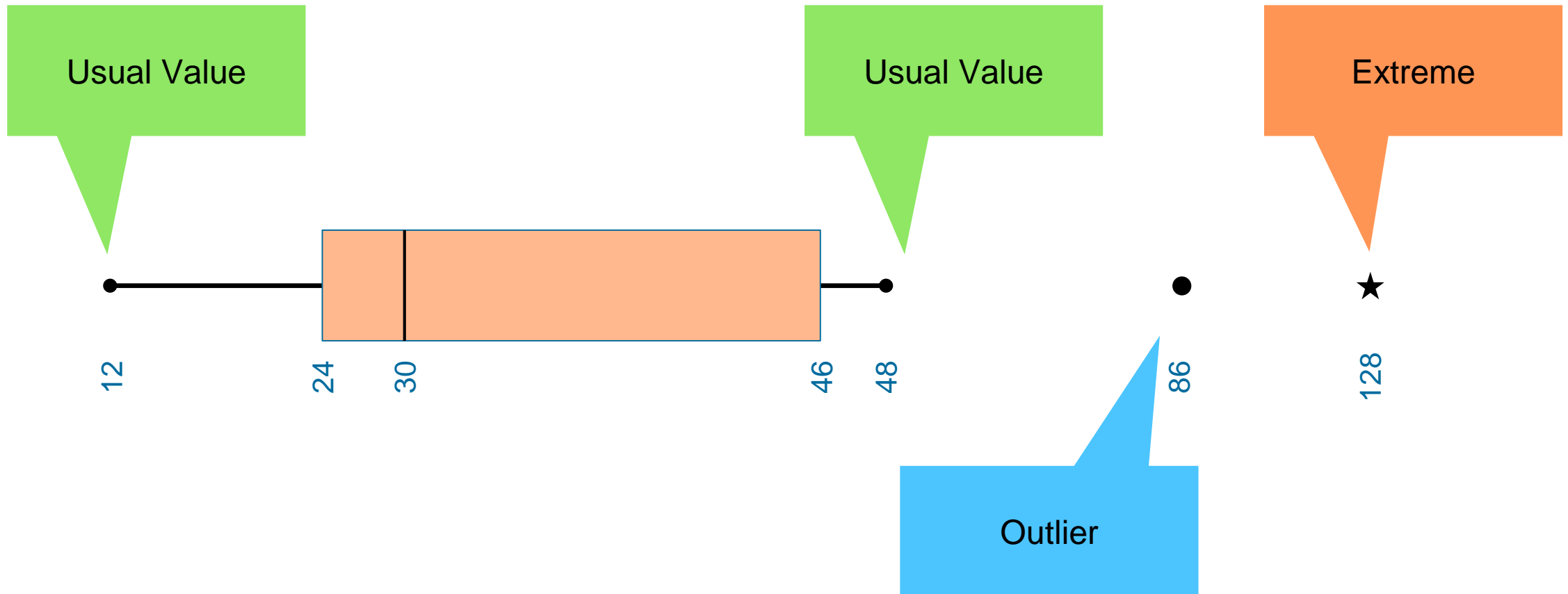
Left Outer Fence	= $Q1 - 3.0 (IQR)$	→	$24 - 3.0 (22)$	= -42
Right Outer Fence	= $Q3 + 3.0 (IQR)$	→	$46 + 3.0 (22)$	= 112

A value that is less than -42 or greater than 112 is an extreme value

12, 20, 20, 24, 25, 26, 27, 30, 32, 34, 36, 46, 48, 86, 128

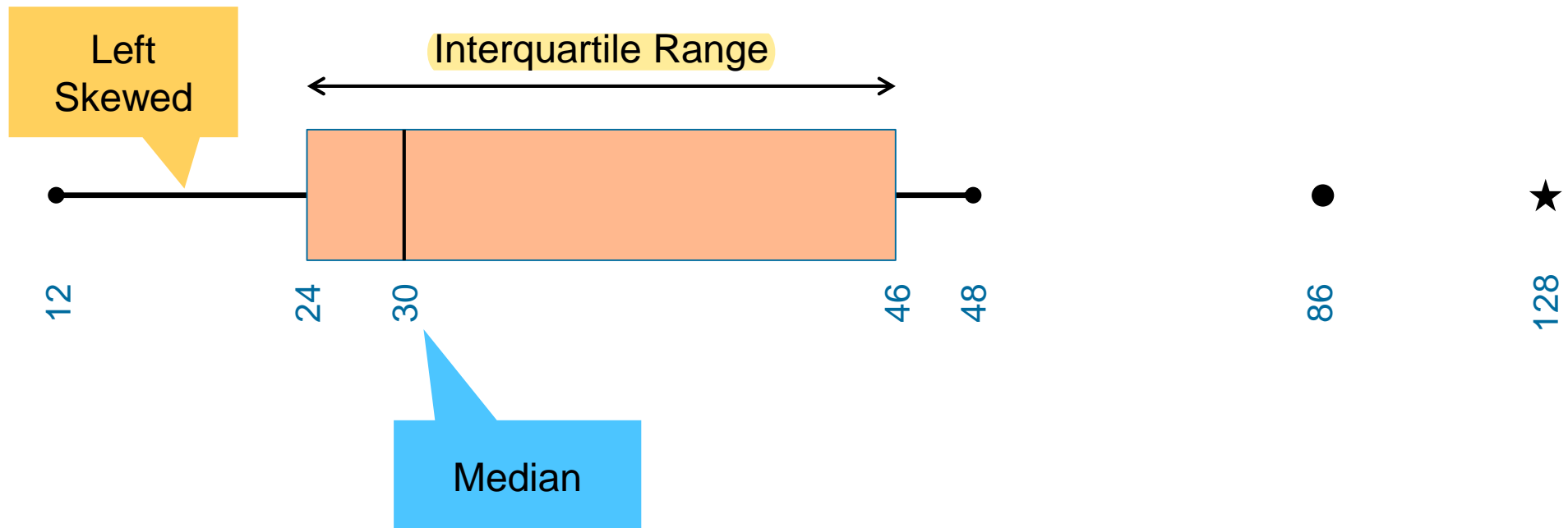
outer ↑  
extreme ↑

## Outlier and extreme value

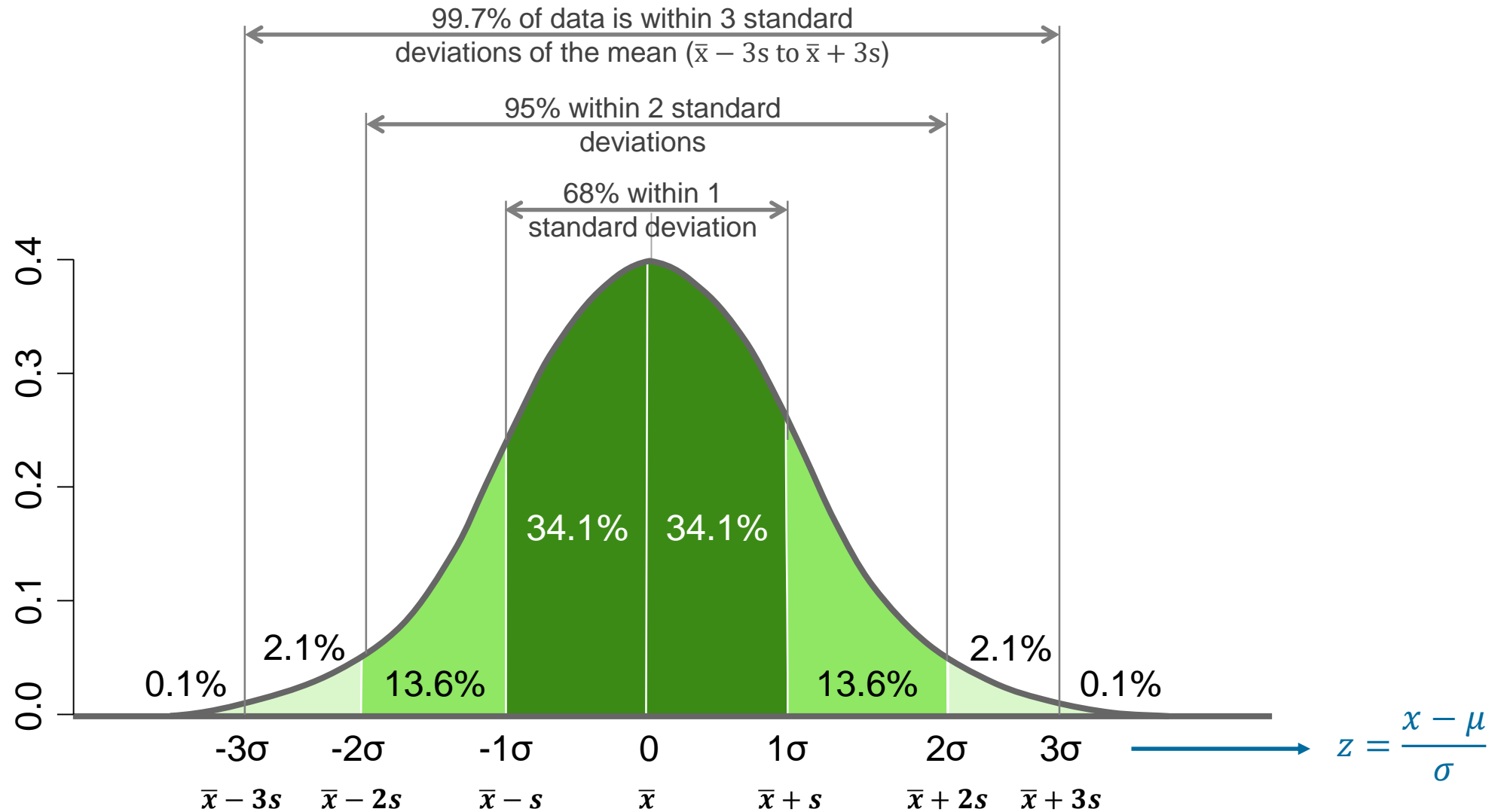




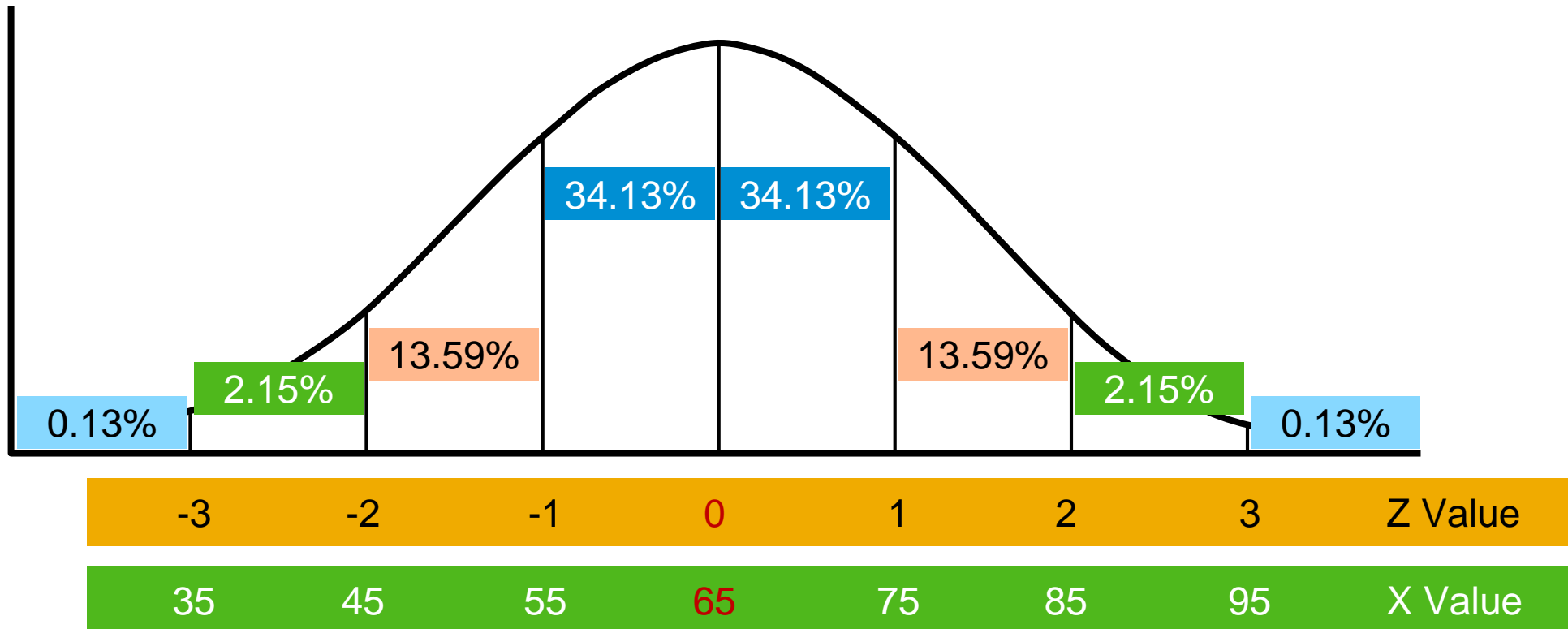
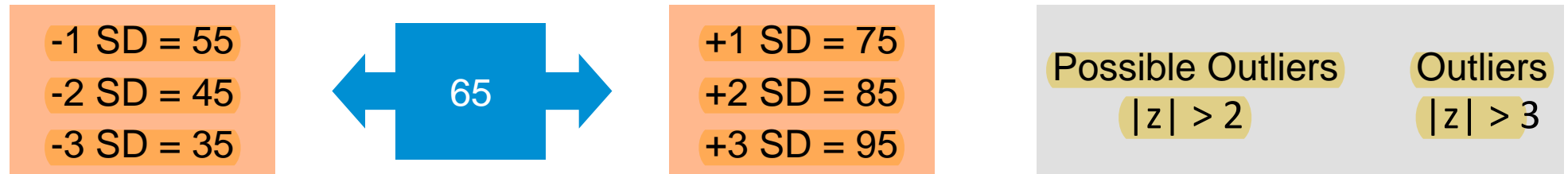
## Summary of box plot features



# The empirical rule



## Rules of thumb for detecting outliers using Z-scores



## Outliers

# Summary

- This lesson has introduced you to some simple but powerful methods to detect outliers.
- You have seen how the interquartile range, box plot, and empirical rule can be used to test for outliers.
- The empirical rule and box plot methods both establish rule-of-thumb limits outside of which a measurement is deemed to be an outlier.
- Usually, the two methods produce similar results. However, the presence of one or more outliers in a dataset can inflate the computed value of the standard deviation. Consequently, it will be less likely that an errant observation would have a Z-score larger than  $|3|$ . In contrast, the values of the quartiles used to calculate the intervals for a box plot are not affected by the presence of outliers.



# Thank you.

**Contact information:**

**open@sap.com**

Follow all of SAP



[www.sap.com/contactsap](http://www.sap.com/contactsap)

© 2019 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platforms, directions, and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See [www.sap.com/copyright](http://www.sap.com/copyright) for additional trademark information and notices.