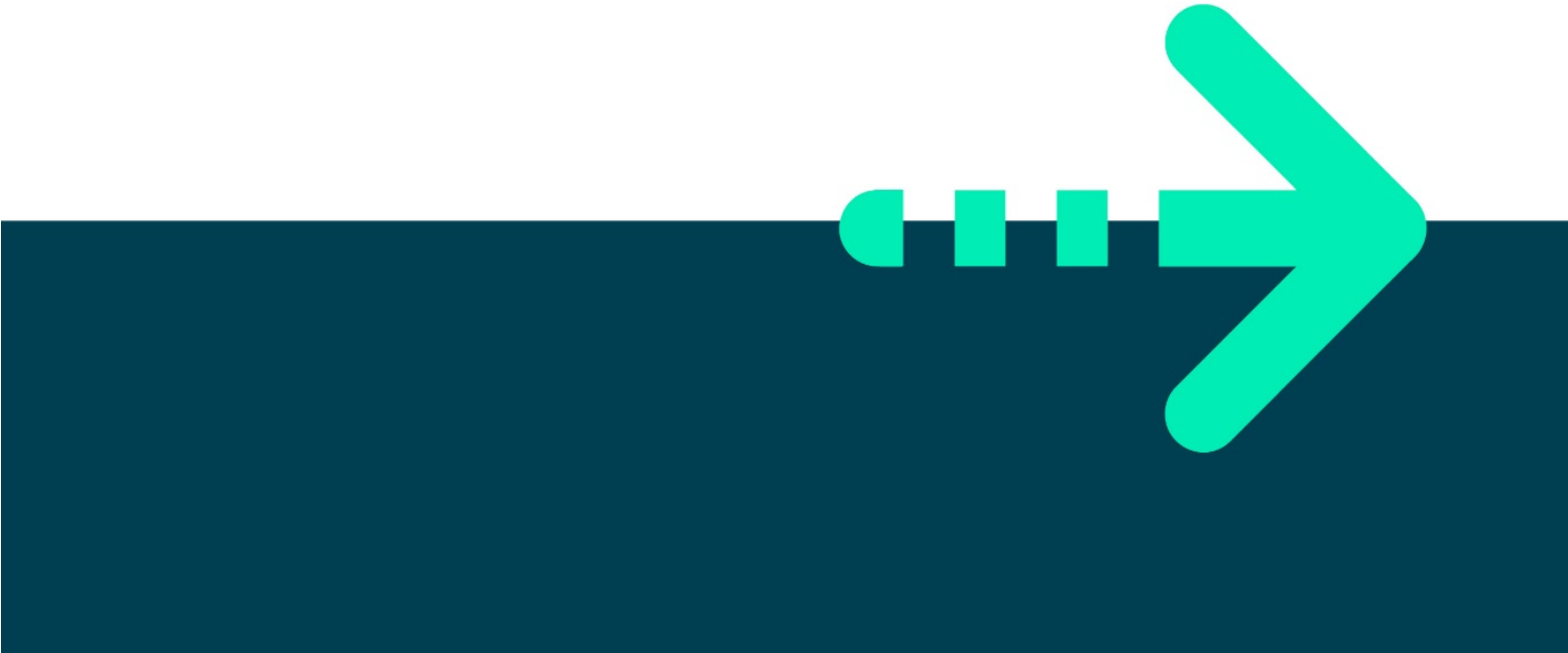
Activity 1: Chart design

Learner Guide



Contents

How to use this workbook

Activity

Alongside this icon you'll find details of the group / individual activity or a point for everyone to discuss.

Useful tool

This icon indicates a technique that will help you put what you learn into practice.

Important idea or concept

Generally, this icon is used to draw your attention to ideas that you need to understand by this point in the course. Let your trainer know if you do not understand or see the relevance of this idea or concept.

Helpful hint

This icon guides you to tips or hints that will help you avoid the common pitfalls or to show you how to increase your effectiveness or efficiency in practising what you have learnt.

Key point

This icon is used to indicate something that practitioners in this field should know. It's likely to be one of the major things to remember from the course, so check you do understand these key points.

Reference material

When we have only touched briefly on a topic, this icon highlights where to look for additional information on the subject. It may also be used to draw your attention to International or National Standards or Web addresses that have interesting collections of information.

Definition

Where a word with a very specific definition (or one that could be described as jargon) is introduced, this will highlight that a definition is provided.

Warning

This icon is used to point out important information that may affect you and your use of the product or service in question.

Introduction

In this activity, you are presented with a dataset that has been visualised poorly. Your **objective** is to implement improvements to the dataset and visualisation. You should try to make your charts understandable to an audience who has no prior knowledge of the dataset. The audience should be able to clearly tell what data they are seeing, where the data came from, and what insights they are supposed to pay attention to.

Dataset and analysis

The data contains the heights and weights of 8,704 individuals. See '**Heights and Weights.xlsx**' workbook on **tblHeightWeight**.

We are given the following metadata as a guide and for further information on the data:

https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overview.aspx?BeginYear=2017

https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/BMX_J.htm

There is no need to read the above metadata documents in detail, but you may need them for clues later.

Bob has analysed the dataset and has produced **tblHeightWeightMeanAll** and the charts in **chart 1** and **chart 2**.

AVERAGEIF

The **AVERAGEIF()** function is used to calculate the average of a range based on a single condition or criteria. The syntax of the function is as follows:

AVERAGEIF(range, criteria, [average_range])

- **range**: The range of cells that you want to apply the criteria to.
- **criteria**: The condition that defines which cells in the range will be included in the average.
- **[average_range]**: (Optional) The actual range of cells that you want to average. If omitted, the function will use the values in the **range** for the average calculation.

AVERAGEIF() function has been applied in **tblHeightWeightMeanAll** to calculate the average height and weight of each gender.

The formula in table for the average height reads as follows:

=AVERAGEIF(tblHeightWeight[Gender], [@Gender], tblHeightWeight[Height])

- **tblHeightWeight[Gender]**: This is the range of cells containing the criteria. In this case, it refers to the 'Gender' column in a table named 'tblHeightWeight.'
- **[@Gender]**: This is a structured table reference that refers to the current row's value in the 'Gender' column. The **@** symbol in Excel tables represents the current row. So, **[@Gender]** refers to the gender value in the current row.
- **tblHeightWeight[Height]**: This is the range of values to be averaged. It refers to the 'Height' column in the 'tblHeightWeight' table.

AVERAGEIFS

The **AVERAGEIFS()** function is used to calculate the average of a range based on multiple criteria. This function allows you to specify conditions that must be met for the average calculation.

The syntax of the AVERAGEIFS function is as follows:

AVERAGEIFS(average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

- **average_range**: This is the range of cells that you want to average.
- **criteria_range1, criteria_range2, ...**: These are the ranges that contain the criteria.
- **criteria1, criteria2, ...**: These are the conditions or criteria that must be met for the corresponding criteria_range.

Here's a simple example:

Suppose you have a dataset with sales data, and you want to calculate the average sales for a specific **product (criteria1)** in a specific **region (criteria2)**, the formula will read as follows:

=AVERAGEIFS(tblsales[sales], tblsales[products], "Product_A", tblsales[regions], "North")

Here's what each part of the formula means:

- **tblsales[sales]**: This is the range of values (sales data) that you want to average.
- **tblsales[products]**: This is the first range of criteria. It represents the products.

- **"Product_A"**: This is the first condition for the product criteria. The formula is calculating the average for sales where the product is "Product_A".
- **tblsales[regions]**: This is the second range of criteria. It represents the regions.
- **"North"**: This is the second condition for the region criteria. The formula is further filtering the average calculation to only include sales data where the product is "Product_A", and the region is "North".

So, the entire formula is calculating the average of sales (**tblsales[sales]**) only for those rows where the product (**tblsales[products]**) is "Product_A", and the region (**tblsales[regions]**) is "North". The **AVERAGEIFS** function is filtering the data based on multiple criteria before computing the average.

You can add more criteria pairs as needed, and the AVERAGEIFS function will consider all specified conditions.

You will be applying AVERAGEIFS function in one of the upcoming tasks.

Task 1

Independent activity:

Identify errors and inconsistencies

Examine the charts, **chart 1** and **chart 2**. Identify at least five things that are unclear or misleading, or that you feel could be improved.

Write your responses on a word document.

Task 2

Independent activity:

Gender and units encoding

Examine tblHeightWeight. By consulting the supplied metadata links below, determine the encoding of gender and the units of height and weight.

https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/DEMO_J.htm

https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/BMX_J.htm

Gender: 1 = ?, 2 = ?

Units of height

Units of weight

Task 3

Spend a couple of minutes and study **tblHeightWeight**, paying particular attention to the **AgeYrs, AgeMonths**, and **Height** columns. Do you think Bob the analyst should have filtered out any erroneous data before computing the average height and/or weight? If so, what filtering would you suggest? Explain your reasons.

Task 4a

Independent activity:

AVERAGEIFS function and chart creation

Create an improved version of **chart 1** to show the mean height and weight by gender. Create two separate charts: one chart showing the male and female height and a second showing the male and female weight.

In **tblHeightWeightMeanAll** and **chart 1**, Bob computed the mean height and weight of all males vs. all females. However, you may wish to average over a **filtered** or **unfiltered** version of the data table as per your answer to task 3.

One way of achieving this, is with the formula function **AVERAGEIFS()**. Please ensure you understand AVERAGEIF and AVERAGEIFS before you attempt this task. See explanations in pages 4 and 5, or your tutor should give a demonstration.

Use **tblHeightWeightMeanFilter** to calculate the average height and weight.

Helpful hint

Note: Criteria such as '> 5' (greater than 5) and '<= 10' (less than or equal to 10) are valid in AVERAGEIFS().

When designing your chart, consider the following questions:

- Can the audience tell what data they are seeing and where the data came from?
- Is the chart appropriately labelled?
- Is the chart legible to individuals with atypical colour vision or visual impairment?
- Are numeric values quoted to a reasonable number of significant figures?
- Does the chart accurately portray what its title suggests? Is the chart accidentally misleading?

Display your charts below the **tblHeightWeightMeanFilter** table on your **Heights and Weights xlsx spreadsheet**.

Task 4b

Modify your formulas in **tblHeightWeightMeanFilter** for the 'Mean height' and 'Mean weight' columns in the table above so that they filter on one additional criterion. The participant's age in years must be greater than or equal to the value specified in the cell below.

Set the minimum age to 0. Verify that the average height and weight for males and females are the same as they were before.

Set the minimum age to 18. Observe what happens to the mean height, the mean weight, and the gap between male and female {height, weight}. Is this what you expected?

Task 5

Independent activity:

Pivot table and pivot chart

Plot the average height of males and females vs. age. Use any type of chart you like. Both male and female heights should be plotted on the same chart as separate series. Filter the dataset as you deem appropriate. Employ the principles of good data visualisation.

Create a pivot table and pivot chart.

To create a pivot table, click any cell inside **'tblHeightWeight'** and go to **Ribbon > Table Design > Tools group > Summarize with PivotTable**.

To create a pivot chart, click any cell inside your pivot table and go to **Ribbon > PivotTable Analyze > PivotChart**.