

Introduction to Power BI

Contents

Setup	2
1 Introduction to Power BI	6
1.1 Overview of Power BI	6
1.2 The parts of Power BI	6
1.3 Use of Power BI and roles	6
1.4 Power BI flow	7
1.5 Use Power BI	7
1.6 Building blocks of Power BI	8
1.7 Power BI Services	10
1.8 Power BI licence	12
2 Importing Dataset and Modelling	13
2.1 Importing data into Power BI Desktop	13
2.2 Dealing with errors	15
2.3 ‘Applied steps’ in modelling data	17
3 Plotting	20
3.1 Visualization Panel	20
3.2 Plot modifications	28
3.3 Pages	30
3.4 Explore Marketplace for other visualizations	30
3.5 Filters	34
3.6 Edit interactions	37
4 Saving and Exporting	39
4.1 Saving as pbix files	39
4.2 Exporting and publishing report	40
5 Next steps	45
5.1 Deepen your understanding	45
5.2 Join the community	45

Setup



These are the course notes for the “Introduction to Power BI” course given by the Monash Bioinformatics Platform¹ for the Monash Data Fluency² initiative. Our teaching style is based on the style of The Carpentries³.

- PDF version for printing⁴
- ZIP of data files used in this workshop⁵

During this workshop we will be using Power BI Desktop installed on your computer. There are several ways to download Power BI Desktop, depending on which system you use.

1. Windows

- **Windows Store**

The first option is to visit Windows Store to get the Power BI Desktop app⁶ and install it. Note that the system requirements is Windows 10 version 14393.0 or higher.

¹<https://www.monash.edu/researchinfrastructure/bioinformatics>

²<https://monashdatafluency.github.io/>

³<https://carpentries.org/>

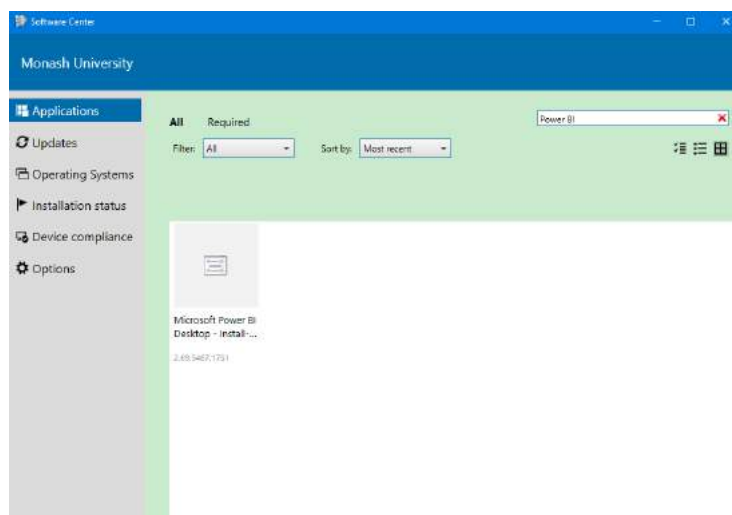
⁴https://monashdatafluency.github.io/Power_BI/powerbi-intro.pdf

⁵https://monashdatafluency.github.io/Power_BI/powerbi-files.zip

⁶<https://www.microsoft.com/en-au/p/power-bi-desktop/9ntxr16hnw1t?activetab=pivot:overviewtab>



If you are using a Monash machine, try searching “**Power BI**” in “Software Center”⁷ and install it. For questions concerning installation on a Monash machine, you may contact eSolutions⁸ for help.



- **Power BI website**

You can also download Power BI Desktop from the Power BI website⁹ and install it as an application on your computer.

⁷<https://www.monash.edu/esolutions/software/install-software-on-monash-devices>

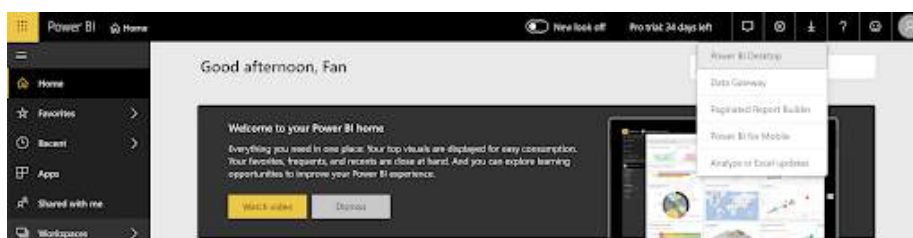
⁸<https://www.monash.edu/esolutions/contact>

⁹<http://go.microsoft.com/fwlink/?LinkID=521662>

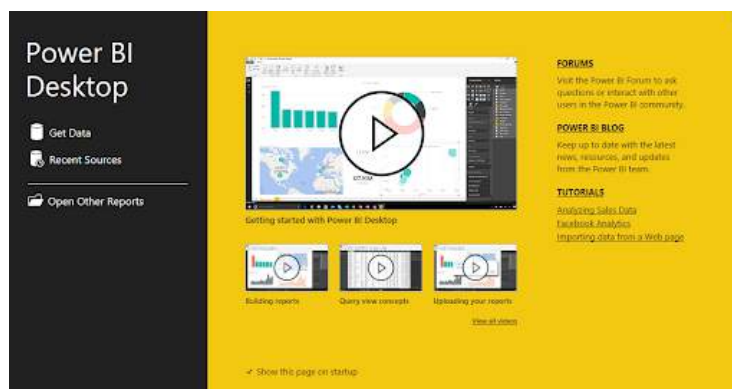


- **Power BI service**

Or if you prefer Power BI Service¹⁰, you can get it by clicking the “**Download**” button in the upper right and selecting “**Power BI Desktop**”. To use Power BI service, you may need to sign in¹¹ your Microsoft account.



After the setup process, you will be able to see the following Start Screen.



2. MacOS

Power BI Desktop is not available on Macs. However, you can run a Windows session on your Mac via BootCamp¹² or using a similar application.

¹⁰<https://app.powerbi.com/home>

¹¹<https://powerbi.microsoft.com/en-us/landing/signin/>

¹²<https://support.apple.com/en-au/HT201468>

3. Linux

Currently it is not possible to install Power BI desktop on Linux. You can choose to install it on a Windows Virtual Machine like VirtualBox¹³ or using a similar application.

After installing Power BI Desktop, you can sign up for Power BI using your Monash account here¹⁴. By signing in the Power BI Desktop, you will be able to save your work and later publish it to the Power BI service.

Data

Download the data file `gap-minder.csv`¹⁵ for the course.

Source code

This book was created in R using the `rmarkdown` and `bookdown` packages.

- GitHub page¹⁶

Authors and copyright

This course is developed for the Monash Data Fluency Team.



This work is licensed under a CC BY-4: Creative Commons Attribution 4.0 International License¹⁷. The attribution is “Monash Bioinformatics Platform” if copying or modifying these notes.

Data files are derived from Gapminder, which has a CC BY-4 license. The attribution is “Free data from www.gapminder.org”. Refer to the Gapminder site¹⁸ for the original form of the data if using it for other uses.

¹³<https://www.virtualbox.org/wiki/Downloads>

¹⁴https://signup.microsoft.com/signup?sku=a403ebcc-fae0-4ca2-8c8c-7a907fd6c235&email=&ru=https%3A%2F%2Fapp.powerbi.com%3Fpbi_source%3Dweb%26redirectedFromSignup%3D1%26noSignUpCheck%3D1

¹⁵https://monashdatafluency.github.io/Power_BI/powerbi-files.zip

¹⁶https://github.com/MonashDataFluency/Power_BI

¹⁷<http://creativecommons.org/licenses/by/4.0/>

¹⁸<https://www.gapminder.org>

Chapter 1

Introduction to Power BI

1.1 Overview of Power BI

Microsoft Power BI is a collection of apps, software services and connectors that come together to turn unrelated data into visually impressive and interactive insights. Power BI can work with simple data sources like Microsoft Excel and complicated ones like cloud-based or on-premises hybrid Data warehouses. Power BI has the capabilities to easily connect to your data sources, visualise and share and publish your findings with anyone and everyone.

Power BI is simple and fast enough to connect to an Excel workbook or a local database. It can also be robust and enterprise-grade, ready for extensive modeling and real time analytics. This means it can be used in a variety of environments from a personal report and visualisation tool to the analytics and decision engine behind group projects, divisions, or entire corporations.

As Power BI is a Microsoft product and has built in connections to Excel, there are many functions that will be familiar to an Excel user.

1.2 The parts of Power BI

Power BI constitutes of a Microsoft Windows desktop application called Power BI Desktop, an online SaaS (Software as a Service) called Power BI Service and a mobile Power BI apps that can be accessed from Windows phones and tablets, and also available on Apple iOS and Google Android devices.

These three elements— **Desktop**, the **Service**, and **Mobile** apps - are the backbone of the Power BI system and lets users create, share and consume the actionable insights in the most effective way.

1.3 Use of Power BI and roles

The use of Power BI could depend on the role that you are in. For example: if you are the stakeholder of a project, then you might want to use **Power BI**

Service or the Mobile **app** to have a glance at how the business is performing. But on the other hand, if you are a developer, you would be using **Power BI Desktop** extensively to publish Power BI desktop reports to the Power BI Service.

In the upcoming modules we would be discussing about these three components - **Desktop**, **Service** and **Mobile** apps - in more detail.

1.4 Power BI flow

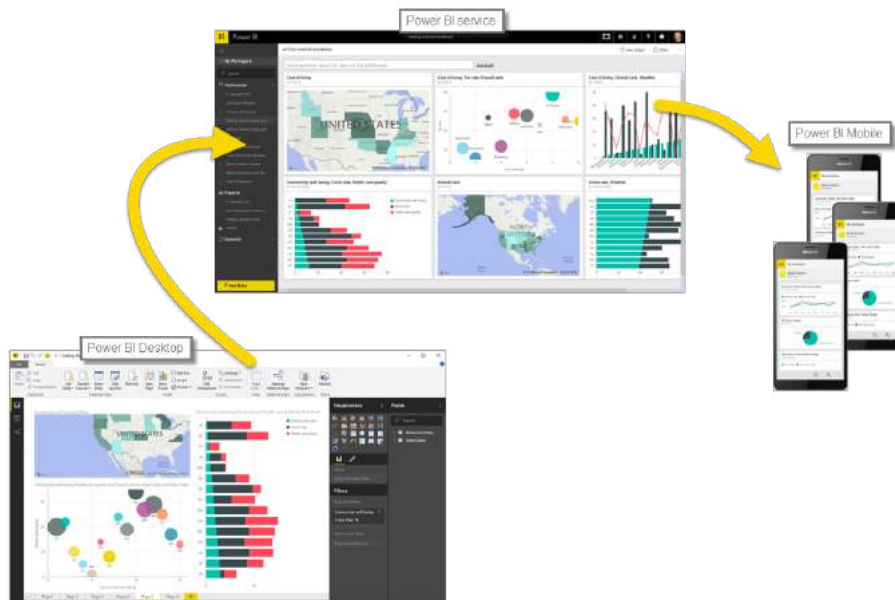
Generally, the flow starts at the Power BI Desktop, where a report is created. This created report can be published to the Power BI Service and finally shared so that the users can use it from the Mobile apps.

This is the most common approach for sharing reports. There are other approaches but we will stick to this flow for this entire tutorial to help learn the different aspects of Power BI.

1.5 Use Power BI

The **common** flow of activity in Power BI looks like this:

1. Bring data into Power BI Desktop, and create a report.
2. Publish to the Power BI service, where you can create new visualizations or build dashboards.
3. Share dashboards with others, especially people who are on the go.
4. View and interact with shared dashboards and reports in Power BI Mobile apps.



Depending on the user role, the user might spend most of the time in one of the three components than the other.

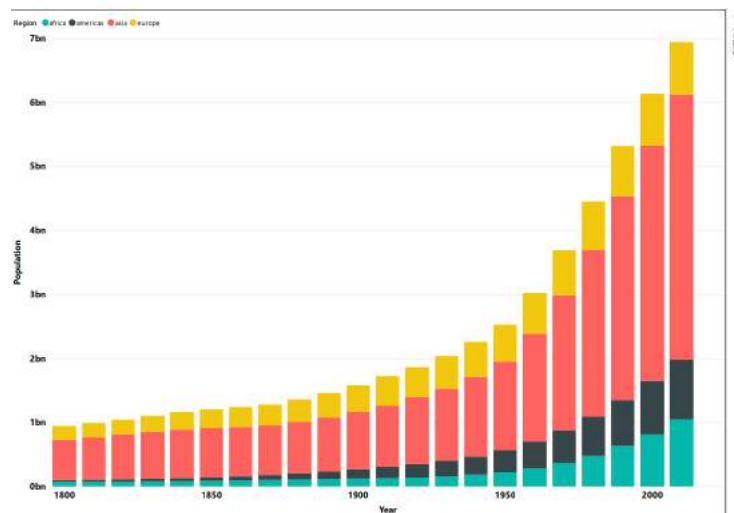
1.6 Building blocks of Power BI

The basic building blocks in Power BI are:

- Visualizations
- Datasets
- Reports
- Dashboards
- Tiles

1.6.1 Visualizations

A visualization is a representation of data in a visual format. It could be a line chart, a bar graph, a color coded map or any visual way to present the data.



Visualizations can be a simple number representing a significant calculation or it could be more complex like multiple charts showing the proportion of users participating in a survey. The main idea of visualisation is to show the data in a way that tells the story that is lying underneath it. Like the saying goes: a picture says a thousand words.

1.6.2 Datasets

A **dataset** is a collection of data that Power BI uses to create its visualizations. You can have a simple dataset that's based on a single table from a Microsoft Excel workbook, similar to what's shown in the following image.

name	year	population	gdp_per_cap	life_exp	region	cont	glt	lat	long	income2017
1 Afghanistan	2000	22000000	300	53.21	asia	ASACZ	70.0	33	69	low
2 Albania	2000	4000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
3 Algeria	2000	29000000	700	70.80	afrika	AFACZ	70.0	30	8	upper_mid
4 Angola	2000	10700000	600	49.00	afrika	AFACZ	15.0	-12	15	low
5 Argentina	2000	36000000	400	74.80	afrika	AFACZ	70.0	33	69	low
6 Azerbaijan and Thebaids	2000	7000000	700	70.80	europa	EUACZ	20.0	39	20	upper_mid
7 Bangladesh	2000	120000000	400	69.4	asia	ASACZ	70.0	33	69	low
8 Australia	2000	20000000	300	74.80	afrika	AFACZ	70.0	33	69	low
9 Austria	2000	8000000	400	74.80	europa	EUACZ	20.0	39	20	upper_mid
10 Bahrain	2000	1000000	1000	74.80	afrika	AFACZ	70.0	33	69	low
11 Belgium	2000	10000000	400	74.80	europa	EUACZ	20.0	39	20	upper_mid
12 Benin	2000	10000000	400	53.21	asia	ASACZ	70.0	33	69	low
13 Bolivia	2000	9000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
14 Botswana	2000	2000000	2000	69.4	europa	EUACZ	20.0	39	20	upper_mid
15 Brazil	2000	170000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
16 Bulgaria	2000	8000000	800	74.80	europa	EUACZ	20.0	39	20	upper_mid
17 Cambodia	2000	12000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
18 Canada	2000	30000000	400	74.80	afrika	AFACZ	70.0	33	69	low
19 Chile	2000	15000000	400	74.80	afrika	AFACZ	70.0	33	69	low
20 China	2000	1200000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
21 Colombia	2000	40000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
22 Costa Rica	2000	4000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
23 Croatia	2000	4000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
24 Cuba	2000	11000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
25 Cyprus	2000	800000	800	74.80	europa	EUACZ	20.0	39	20	upper_mid
26 Czechia	2000	10000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
27 Denmark	2000	5000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid
28 Dominican Republic	2000	7000000	400	69.4	europa	EUACZ	20.0	39	20	upper_mid

Dataset can also be a combination of many different sources, which can be filtered using Power BI and combined into one to use.

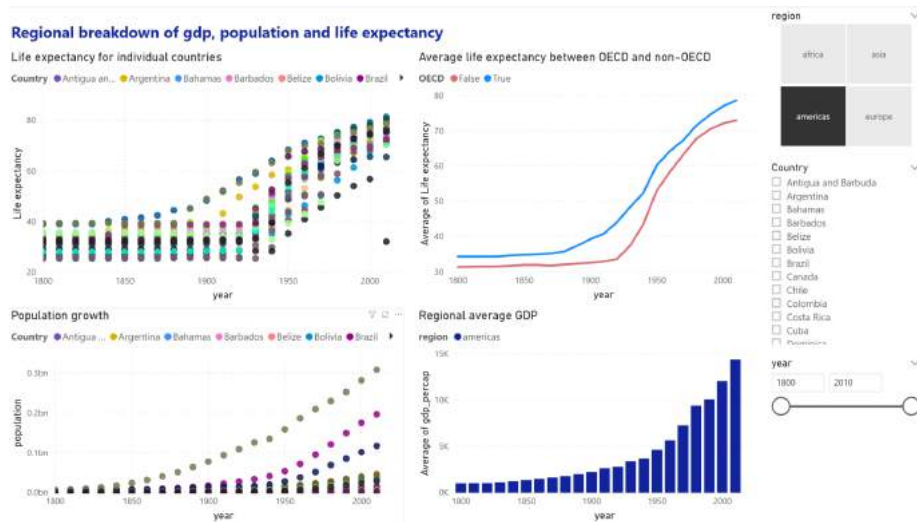
For example: One data source contains countries and locations in the form of latitude and longitude. Another data source contains demographics of these countries like population and GDP. Power BI can combine these two data sources into one dataset which can be used for visualizations.

An important feature of Power BI is the ability to connect to various data sources using its connectors. Whether the data you want is in Excel or a Microsoft SQL Server database, in Azure or Oracle, or in a service like Facebook, Salesforce, or MailChimp, Power BI has built-in data connectors that let you easily connect to that data, filter it if necessary, and bring it into your dataset.

After you have a dataset, you can begin creating visualizations that show different portions of it in different ways, and gain insights based on what you see. That is where reports come in.

1.6.3 Reports

In Power BI, a **Report** is a collection of visualizations that appear together on one or more pages. A report in Power BI is a collection of items that are related to each other. We will be working with the gapminder data to create the report below that looks at the GDP, population and life expectancy by global regions.



Reports let us create and structure visualizations on pages based on the way the we want to tell the story.

1.6.4 Dashboards

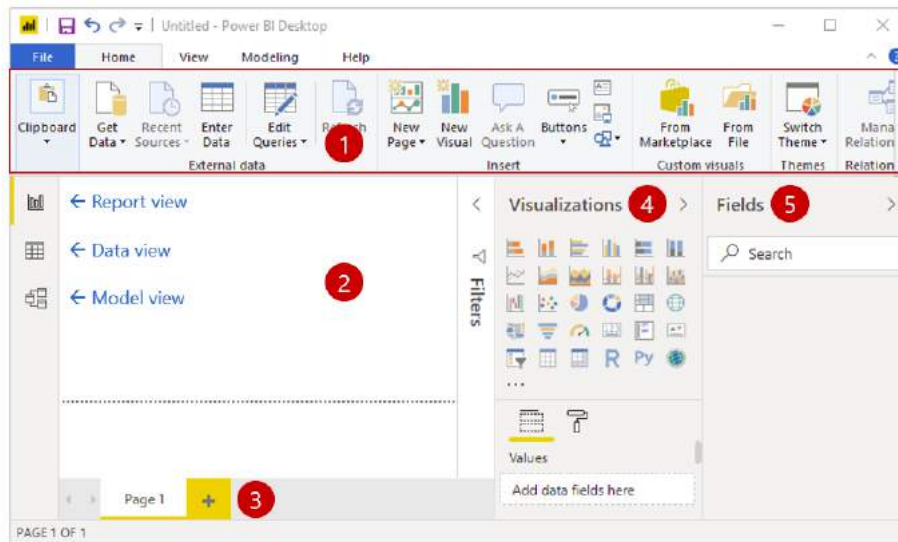
A Power BI dashboard is a collection of visuals from a single page that you can share with others. Often it is a selected group of visuals that provide quick insight into the data or story you are trying to present.

A dashboard must fit on a single page, often called a canvas (the canvas is the blank backdrop in Power BI Desktop or the service, where you put visualizations). Think of it like the canvas that an artist or painter uses — a workspace where you create, combine, and rework interesting and compelling visuals. You can share dashboards with other users or groups, who can then interact with your dashboards when they're in the Power BI service or on their mobile device.

1.7 Power BI Services

1.7.1 Overview of Power BI Desktop

Power BI Desktop is a free application for PCs that lets you gather, transform, and visualize your data. In this module, you'll learn how to find and collect data from different sources and how to clean or transform it. You'll also learn tricks to make data-gathering easier. Power BI Desktop and the Power BI Service work together. You can create your reports and dashboards in Power BI Desktop, and then publish them to the Power BI Service for others to consume.



1. **Ribbon** - Displays common tasks that are associated with reports and visualizations.
2. **Report view, or canvas** - Where visualizations are created and arranged. You can switch between **Report**, **Data**, and **Model** views by selecting the icons in the left column.
3. **Pages tab** - Located along the bottom of the page, this area is where you would select or add a report page.
4. **Visualizations pane** - Where you can change visualizations, customize colors or axes, apply filters, drag fields, and more.
5. **Fields pane** - Where query elements and filters can be dragged onto the **Report** view or dragged to the **Filters** area of the Visualizations pane.

Working in Power BI desktop and creating visuals will be the focus of today's session.

1.7.2 Power BI Service

Power BI Service is the online component of Power BI where you can publish your dashboards and reports. You can also view other dashboard and reports that have been shared with you. Monash staff members have access to this, all you need to do is log in with your Monash account at Power BI service.¹ We will cover this in more detail later in the workshop. One thing to be aware of is that once the report is published, the report and underlying data will be stored on Microsoft servers and is not private. Be aware of any privacy or confidentiality issues with your data and we suggest using another approach or tool if cannot anonymise your data.

¹<https://powerbi.microsoft.com/en-us/landing/signin/>

1.8 Power BI licence

We will be covering the free version of Power BI desktop in this session. This version allows you to use all the functionalities of the desktop version such as loading, analysing and visualising data. You can create as many visuals and reports as you would like with the free license. However publishing and sharing your reports and visuals online with another person or publically requires a Power BI license. This means that you can still share your report and data files between users and computers through other means, like a shared drive, and only require one user with a Power BI license to publish and share the final report online. Please contact the Monash Business Intelligence team for more information on licensing.

Chapter 2

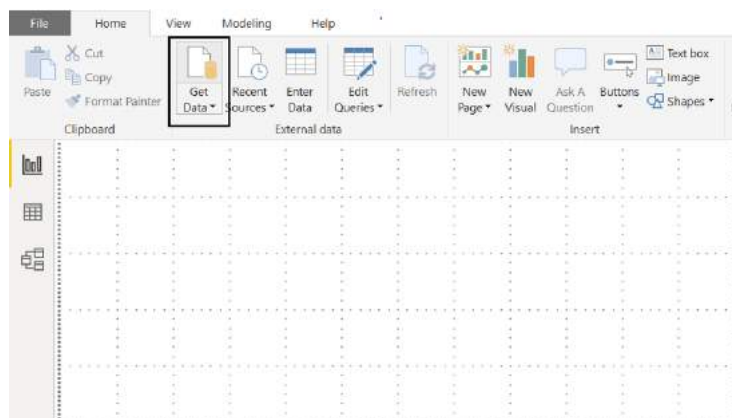
Importing Dataset and Modelling

Power BI can connect to a whole range of data sources from Excel sheets and local databases to several Cloud services. Currently, over 60 different cloud services have specific connectors to help you connect with generic sources through XML, CSV, text, and ODBC. Let us start connecting to one of the data sources. For today we will be working on the `gap_minder_map.csv` file.

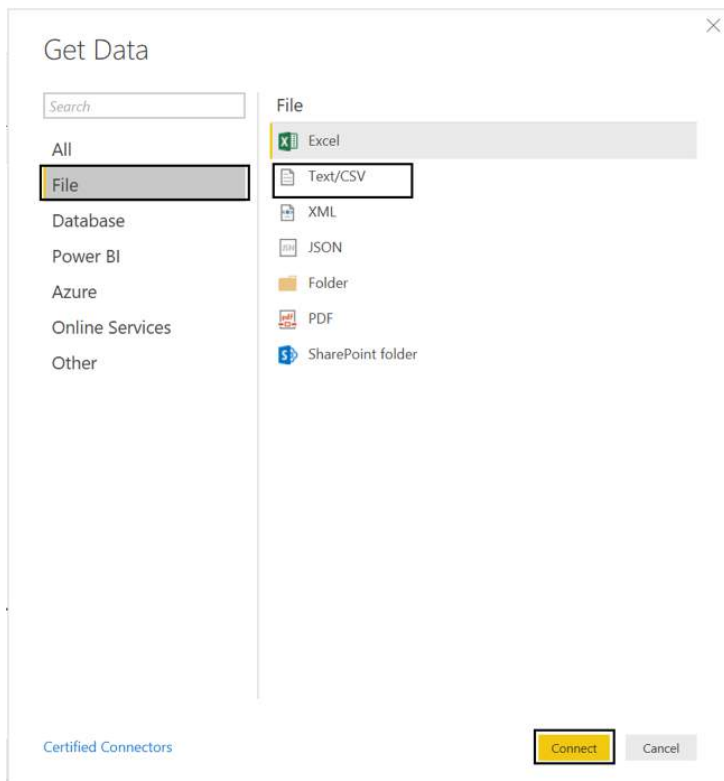
2.1 Importing data into Power BI Desktop

Power BI Desktop has a **Get Data** button from the ribbon on the **Home** tab. In Power BI, there are all sorts of different data sources available. Select a source to establish a connection. Depending on your selection, you will be asked to find the source on your computer or network, or be prompted to sign in to a service to authenticate your request.

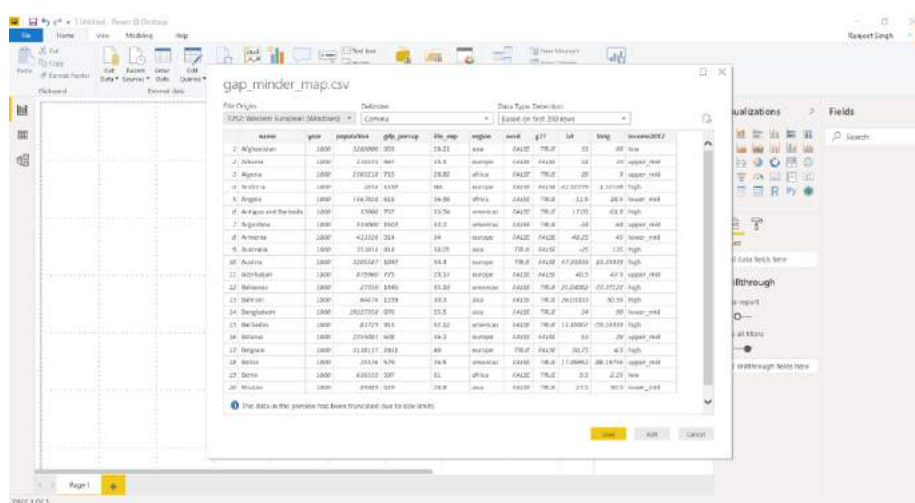
As our first step to import the dataset/file into Power BI, we click on the Get Data icon on the ribbon of Home tab.



Once we select this, we go ahead and select the **CSV** option under the **file** subheading.



Then browse the file and select the necessary CSV file. Press on **Connect** to have a quick preview of the file. Once we click on **Load**, Power BI will successfully import the file. Any errors will then pop up ready to be analysed and fixed.



id	name	year	gap_minder	gap_exp	region
1	1. Afghanistan	1880	2328000	15.4	north
2	2. Albania	1880	413640	15.4	north
3	3. Algeria	1880	2509218	15.4	north
4	4. Argentina	1880	1648	15.4	north
5	5. Austria	1880	1367658	15.4	north
6	6. Australia and Oceania	1880	17080	15.4	north
7	7. Belgium	1880	118380	15.4	north
8	8. Bolivia	1880	413018	15.4	north
9	9. Brazil	1880	1111118	15.4	north
10	10. Canada	1880	1000000	15.4	north
11	11. Chile	1880	817000	15.4	north
12	12. Colombia	1880	17010	15.4	north
13	13. Costa Rica	1880	66476	15.4	north
14	14. Cuba	1880	1367718	15.4	north
15	15. Denmark	1880	81700	15.4	north
16	16. Ecuador	1880	1111118	15.4	north
17	17. Egypt	1880	1111118	15.4	north
18	18. El Salvador	1880	1111118	15.4	north
19	19. Finland	1880	1111118	15.4	north
20	20. France	1880	1111118	15.4	north
21	21. Germany	1880	1111118	15.4	north
22	22. Greece and Turkey	1880	1111118	15.4	north
23	23. Guatemala	1880	1111118	15.4	north
24	24. Haiti	1880	1111118	15.4	north
25	25. Honduras	1880	1111118	15.4	north
26	26. India	1880	1111118	15.4	north
27	27. Indonesia	1880	1111118	15.4	north
28	28. Iran	1880	1111118	15.4	north
29	29. Italy	1880	1111118	15.4	north
30	30. Japan	1880	1111118	15.4	north

If there are any errors in the data load, select **View errors** to check any detected errors. Alternatively, close the pop up and click **Edit queries** to check for any errors. This will query and list the errors in the data.

2.2 Dealing with errors

2.2.1 Checking for automatically detected errors

Once queried, any detected errors will be displayed. Clicking on these errors will provide a brief description on what might have gone wrong. There will be some errors showing up in your data, click on the errors to see what has caused the issue.

2.2.2 Changing data type of column

One of the most common errors is the detection of the datatype of a particular column. While loading the data, Power BI automatically assigns a column type based on the first lines of data. This can cause errors or other problems if the automated data type is not suited for that column. You may need to change data types of the columns to a relevant type for your data model.

Any changes to the data needs to be done under the **Data file** listed under **Other Queries**.

The **life_exp** column is displaying as text data. This is a problem as we may want to do numerical calculations with this data. We should change this to a numerical data type.

- Select the **gap-minder** file.
- Right click on top of the column and select **Change type**.
- Change to **Decimal Number**.

You can replace a current step in the query with **Replace current step** or add a new step to the query with **Add new step**.

- Click **Add new step**.

Once necessary changes are made, it is important to refresh and check if the change was applied. We can do this by clicking on the **Refresh Preview** button on the ribbon.

You will notice that this has created errors due to several cells containing N/A.

2.2.3 Replacing values

Data can have erroneous and missing values for a number of reasons. Missing data is represented as `null` in the data. You will often need to deal with such values and fix or remove them.

You can view the data in the **Data** section on the left sidebar. By applying a filter to a particular column, the null values can be analysed. To replace any null values, go to the **Query Editor** and use the **Replace values** option in the **Transform** tab. You can also remove error values by right-clicking a column and selecting the **Remove errors** option.

Look in **Errors in gap-minder** in the Query Errors folder. You will notice that all the years for 1820 have a whitespace in it that is causing an error.

- Return to the gap-minder data .
- Right click the **year** column and select **Replace errors**.
- Replace this with 1820.

This will replace all errors with the value 1820. In this case this works as only one value is causing an issue. You can use **Replace values** to find and replace several values.

Challenge 1: Replace missing values

Use the **Remove errors** option on the `life_exp` column to remove the error values caused by the blanks. Use the **Replace values** option on the `gdp_percap` column to replace the blank values.

2.2.4 Final notes on dealing with errors

Replacing error and missing values is not always a direct operation. Most of the time, these values must be carefully analyzed and replacement values need to be computed based on several factors. While this can be done with in Power BI,

this is not a key capability of Power BI. Furthermore these changes do not affect the source data set. These changes are actions that Power BI takes whenever you load in that dataset.

You may have noticed errors in the `OECD` column. If you tried to fix these with a **Replace values** option you will find that this does not work. This is because Power BI has automatically made this column a Boolean column and is intentionally ignore this values as they produce errors. To get around this you would have to change the column to a text column, find and replace the values, and then change the column back to boolean. This are additional steps that Power BI will be made to run everytime you load this data and is inefficient.

Cleaning up the source data as much as possible in the early stages will save you problems in the future, this is true for any data analysis. The best approach is to clean up and eliminate errors from your data before you import into Power BI. This will minimise the amount of actions that Power BI takes and speeds up the loading process. This cleaning process may take several iterations as you work with Power BI and the data and uncover more inaccuracies.

Power BI does ignore blank and error values when it comes to analysis and visualisation but this may not always be the best approach.

Cancel the last two steps in the **Applied steps** by clicking on the red crossmark right next to **Removed Errors** and **Replaced Value**.

2.3 ‘Applied steps’ in modelling data

2.3.1 Renaming columns

Columns from raw data can be difficult to read or meaningless. Renaming the columns in your query to a meaningful name will make it easier for you and your audience to understand your data. This will often save you trouble in the future when it comes to working and presenting the data.

There are two ways to rename the columns in Power BI. Right-clicking on the header of the column gives you a menu of functions that you can do to the column. Select **Rename** to rename the column. You can also click on the column and then click on the **Transform** tab, from here you are presented a variety of transformation functions for the column. From here you can click on **Rename**.

- Right-click on the `name` column header, click on **Rename** and rename the column to `Country`.
- Click on the header for `life_exp`. Click on the **Transform** tab and click rename. Rename the column to `Life expectancy`.

2.3.2 Adding and removing columns

2.3.2.1 Removing columns

Often when dealing with raw data you will find columns that are meaningless or unsuitable for your analysis. You can remove these columns in Query to eliminate clutter and streamline the data set making it easier to work with.

There are two ways to remove columns in Power BI. As in renaming a column, you can right-click the header, which will present a menu with the option to **Remove**. You can also click on the column, click on the **Home** tab and then click **Remove Columns**.

We will not be using the `g77` information.

- Right-click on the `g77` column header, click on **Remove**

2.3.2.2 Adding columns

Just as you would remove unsuitable data from your queries, you may need to add new columns to your data. There are a variety of options in Power BI to add different columns. Click on the **Add Column** tab to see the ways you can add a column.

- Click on `gdp_percap`, click on the **Add Column** tab, click duplicate to create a separate column to run calculations on.

Challenge 2: Add

Perform a simple mathematical operation and add this as a new column. Let's say you want to calculate Total Gross Domestic Products (GDP) per country. As $\text{Total GDP} = \text{Population} \times \text{GDP per capita}$, using the new `gdp_percap` column you created in the previous example, use **Custom Column** to create a new column with Total GDP per country.

2.3.3 Navigating 'applied steps' and undoing

It is important to know what sort of transformations the data has gone through to understand the current state of your data. These steps are listed on the **Applied Steps** section on the **Query Editor**. Sometimes these steps might need to be moved up or down based on the required application. These steps can also be removed by clicking on the red crossmark next to the step. This will undo the action it performs and returns the data to the previous stage.

2.3.4 Refreshing data ready for analysis

Always refresh your data when you are ready to do your analysis and visualisation. Clicking the Refresh option will refresh the data and rerun all your queries. This will show any issues or errors with your data or your queries and ensure that all the data is up to date and ready for analysis.

Once all the required changes are done, select **Close & Apply** going back to the main Power BI Desktop interface.

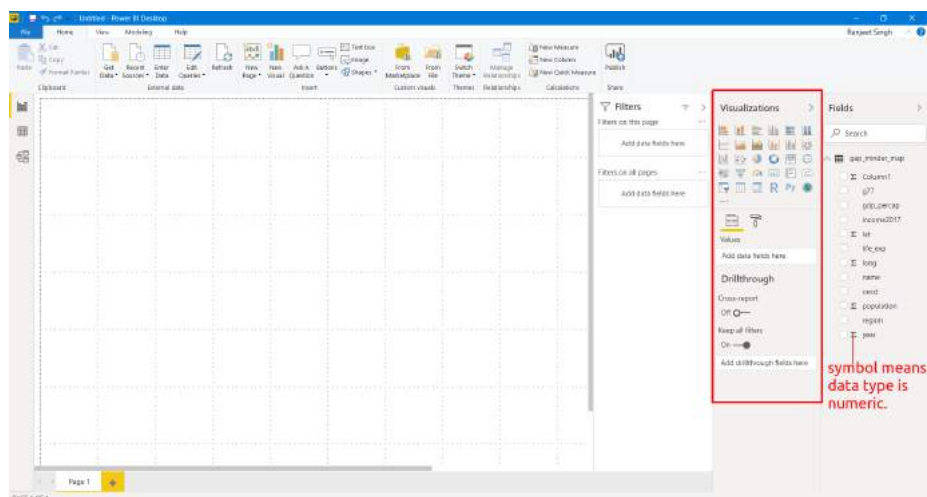
Chapter 3

Plotting

In the previous chapter, we saw how to import data into Power BI. We also did some basic pre-processing of data like cleaning the data, changing the type of data, and dealing with null and error values. In this section, we are going to work with our already processed data.

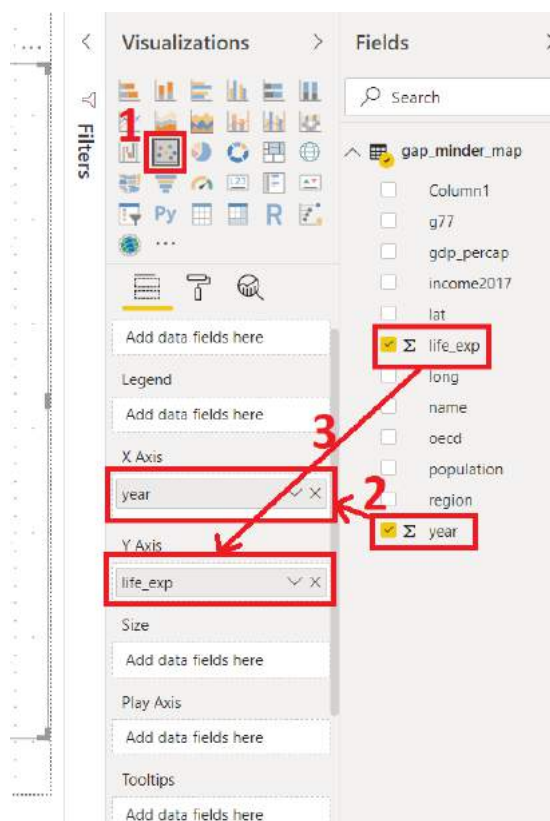
3.1 Visualization Panel

We touched upon the different panels present in the Power BI interface. We will now explore the different panels dedicated to creating and formatting charts in detail. One of the panels is the **Visualization panel** highlighted in the Red box. It contains various types of charts that you can create to visualize your data. Some of the available chart types are - Stacked bar chart, Line chart, Area chart, Scatter plot, Pie chart, etc. You can also import a custom visual from a file or the marketplace if you click on the ... icon. We will explore that in the later section.



3.1.1 Scatter plot

Let's create our first chart, a Scatter plot. The first step to plot any kind of plot in Power BI is to select the plot template/type from the **Visualization panel**. Click on the scatter chart icon in the visualization panel and drag the columns to the respective -Axis field as shown in the below figure.



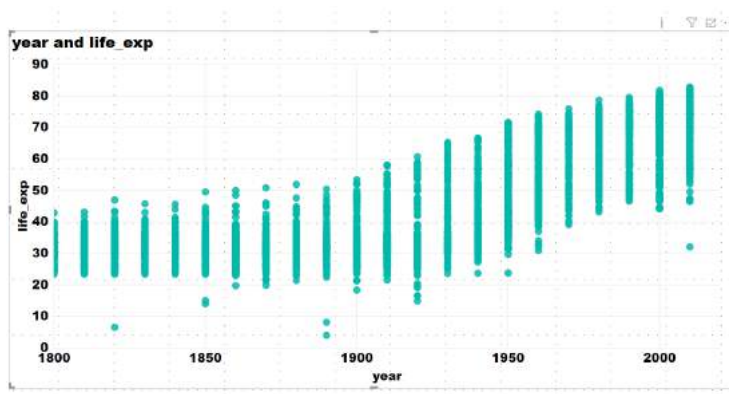
We want to see how life expectancy `life_exp` varies with time `year` in years. After dragging the columns to their respective fields, we get our first scatter plot.

This plot does not seem entirely correct.

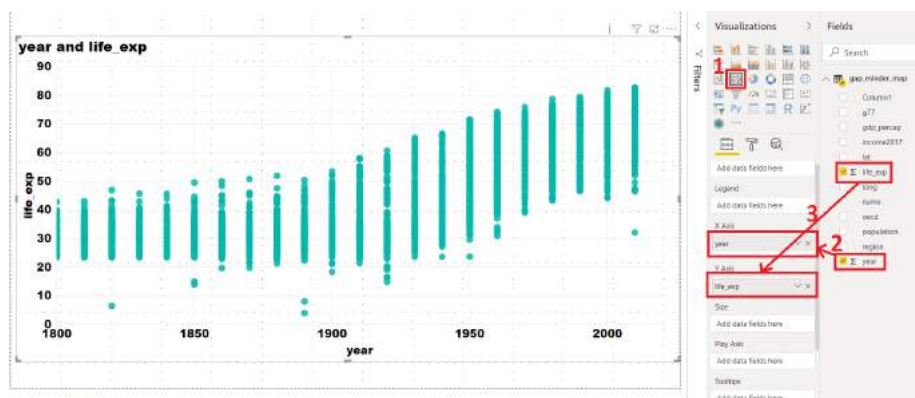
- Click the drop down on both `life_exp` and `year` fields. These fields are being automatically summed.
- Click don't summarize.

PowerBI automatically summarizes numerical data. Be aware and check this when you add new numerical data to visualization fields.

An overall increasing trend can now be seen after year 1900.



The final interface will look like the image below. The plot will be at the left of the **Visualization panel**.



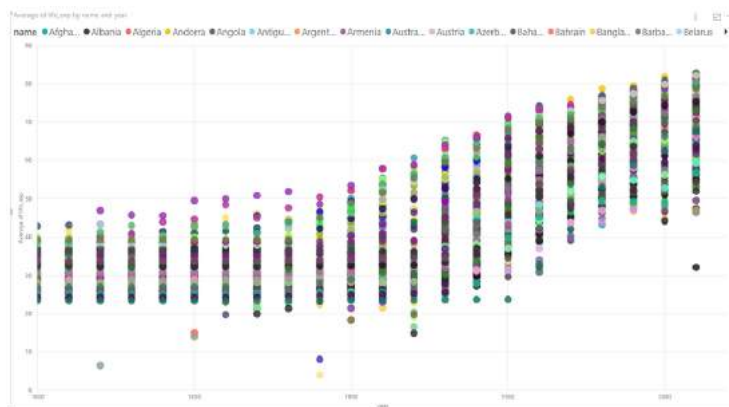
In general, this will be the procedure to create any type of plot in Power BI i.e. select the chart type from the **Visualization panel** and then drag data to the relevant fields.

The above scatter plot is not informative. Let's explore what else we can do with this plot.

X-Axis: year

Y-Axis: life_exp

Legend: country

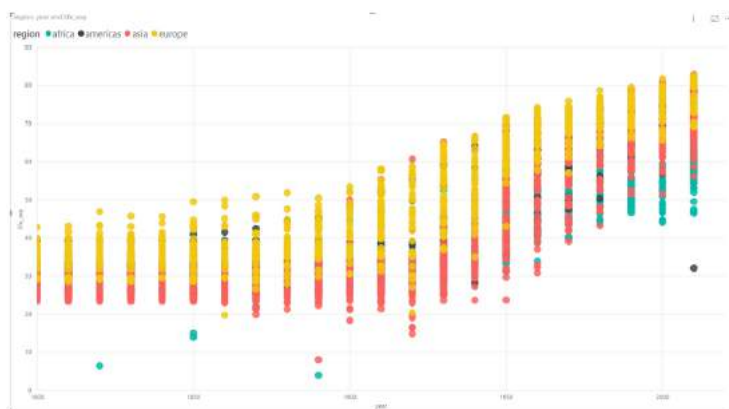


We have colored each dot by the country. This plot shows the trend of life expectancy for **each country** with years. Since there are so many countries in our data, this plot looks messy, unclear and our plot legend “explodes”.

Let’s make this better.

Use the following parameters to create your chart:

X-Axis: `year`
Y-Axis: `life_exp`
Legend: `region`



This plot colors each data point by the region where the country is located. The **summarize** option summarizes the life expectancy `life_exp` for each region.

In the same chart, try the following:

- Size: By `population`
- Report what follows.

Challenge 1 Scatter plot

Create a chart with the following options:

X-Axis: year

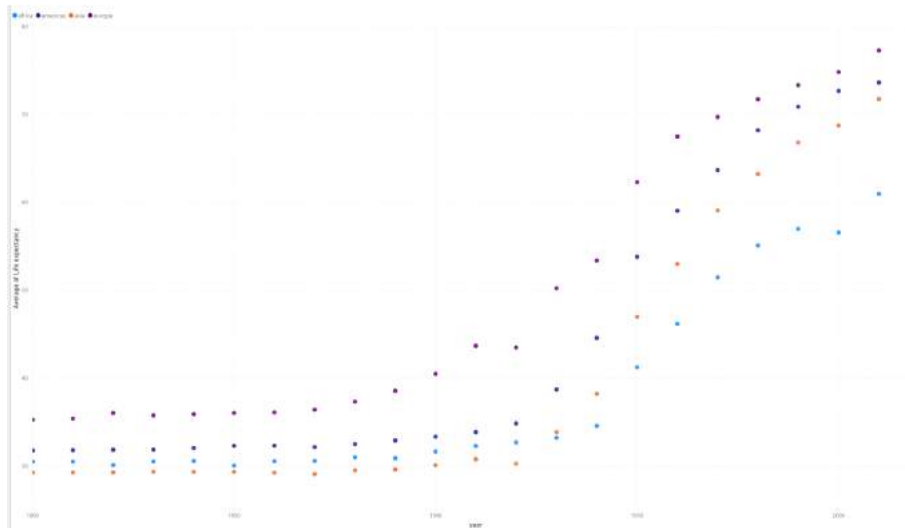
Y-Axis: life expectancy

Legend: By region

Find the **Average** life expectancy of each region by year.

Challenge 1: Solution

This shows the average of **life expectancy** in all countries in a given region, and plots it.



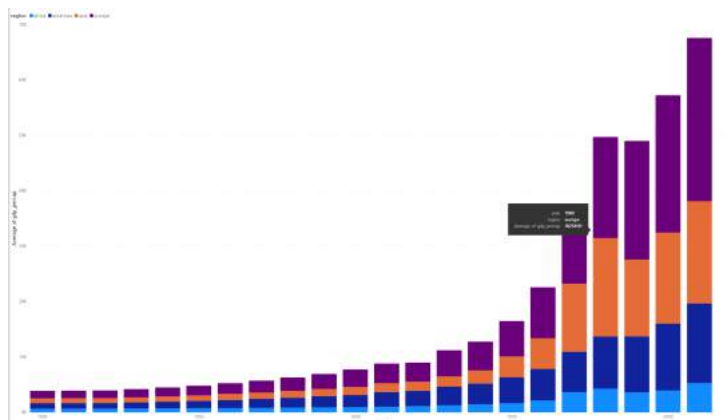
3.1.2 Bar plot

Select the bar chart from the **Visualization panel** and drag the following columns:

Axis: Year

Value: GDP per capita

Legend: By region

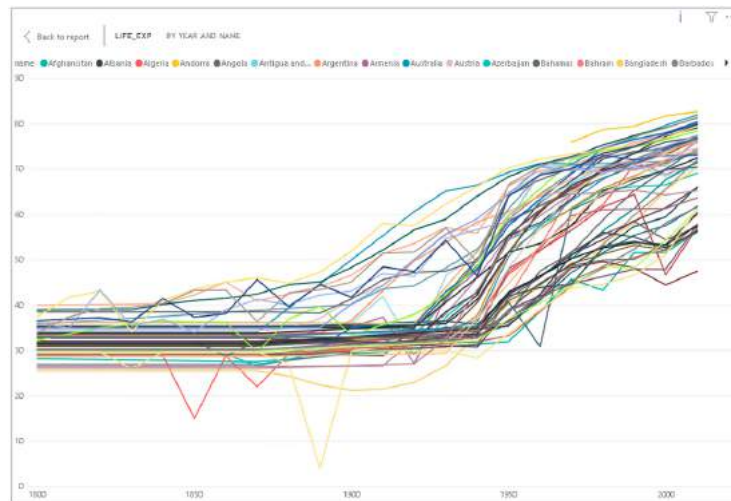


This bar plot shows GDP per capita variation over time (in years) for all 4 regions.

3.1.3 Line chart

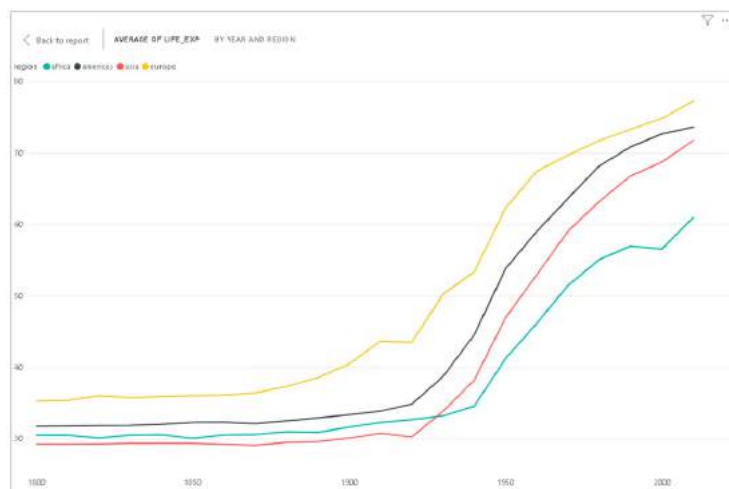
Select a line chart from the **Visualization panel**. Drag and drop the following columns:

Axis: Year
 Values: Life_exp
 Legend: By name



We get differently colored lines for each country. Now try changing the Legend to by region.

Axis: Year
 Values: Average of Life_exp. Select Average
 Legend: By region



Power BI computes average of Life expectancy for countries in each region and shows four average lines corresponding to four regions.

3.1.4 Changing visualizations

As you can see from your line chart, this uses the same data as your scatter plot but presented in a different visual format. Instead of making a new separate visualization to do a different visual style, you can edit and change existing visualizations easily with Power BI.

For the current line chart, select the visual and try selecting different visualizations on the **Visualization panel**. Visualizations with similar structures to your existing visual, like bar charts, will change over seamlessly. A slightly different visualization like a scatter plot might require some testing with the **Fields** in the **Fields** tab.

- Change your line chart to a scatter plot.
- Drag `life_exp` to the y-axis and `Year` to the x-axis

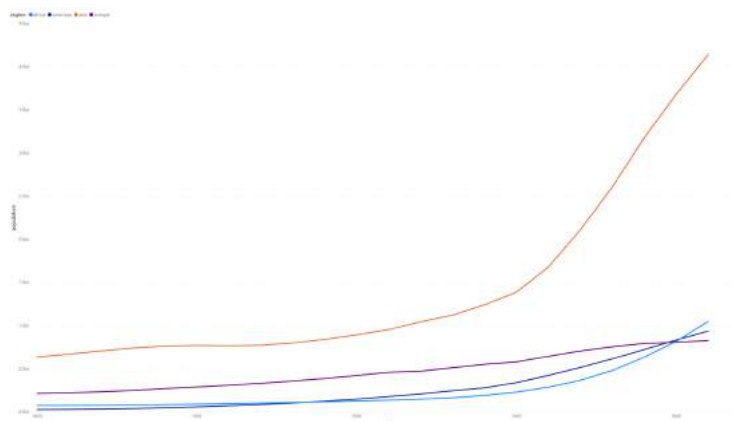
This should now be identical to your first scatter plot.

As you may have noticed, dragging and dropping **Fields** into the **Fields** tab will automatically change the data visualization. You can easily swap or add new fields using the drag and drop to change and discover new ways to visualize your data. You can remove fields by clicking the x next to the name.

Challenge 2: Line chart

As the line chart is showing the same information as an existing visualization, change the chart to show population instead of average life expectancy.

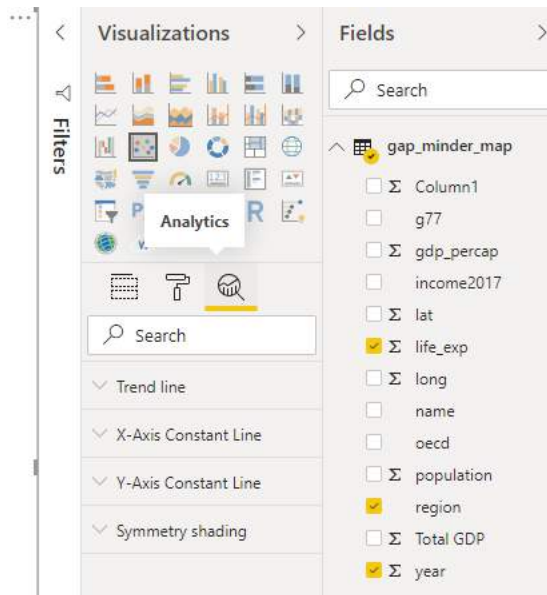
Challenge 2: Solution



3.1.5 Trends and analytics in plots

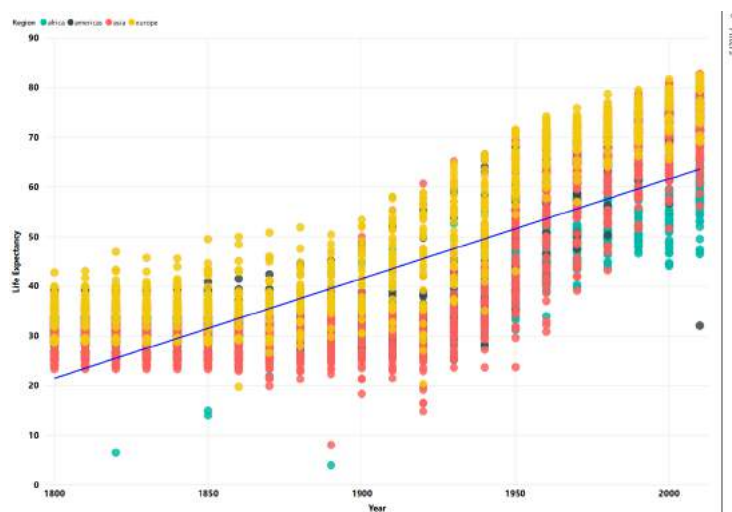
Up till now we have been working in the **Fields** tab of the **Visualizations** panel. There are two other useful tabs, **Format** and **Analytics**. Let's have a look at **Analytics**.

This tab allows you to add different reference lines to your visualization which can help the user determine trends and insights in the data. Click on the scatter plot that you have created. Select the **Analytics** tab.



Here you can add reference lines for various measures. Add a trend line to your chart.

These options will vary depending on the chart that you use.



3.2 Plot modifications

An important aspect of visualizing and telling the story of your data is the design and format of your visualizations. This is where we will use the **Format** tab.

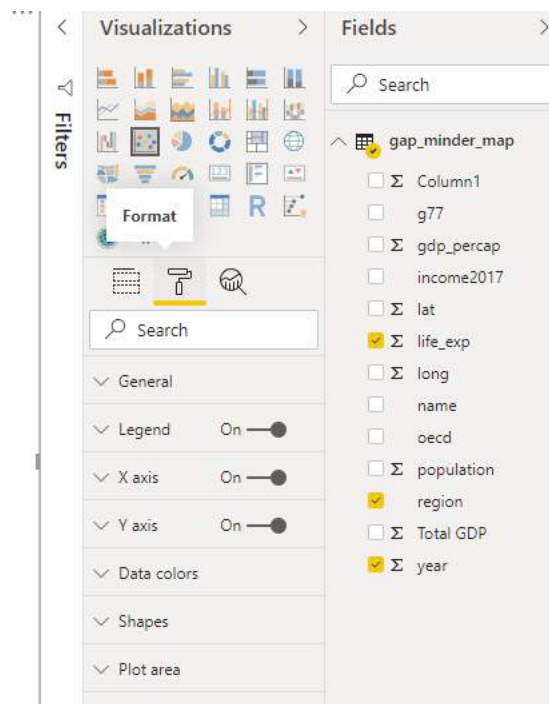
Click on your scatter chart you created for Challenge 2. Click the **Format** tab to see the variety of options available to make your visualization more readable and understandable.

A common issue is the default titles of charts. This will usually use the names of the data columns which is often meaningless to the average user.

- Click on **Title** and change it to a clearer name like 'Regional average life expectancy since 1800'.

You can also edit the legend or axis labels. Let's make the axis labels more clear.

- Change the font for the x and y axis to Segoe(Bold), increase the font size to 12 and color to black.



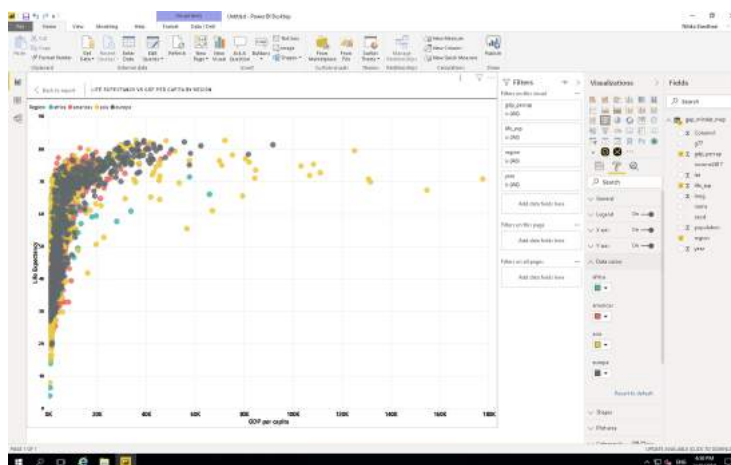
The options available will be different for each visualization. Using these options will greatly increase the engagement and understanding of your visuals. Let's go and change the names of all the other visuals to something more meaningful.

Challenge 3: Scatter Plot

Let's say you think there might be a correlation between GDP per capita and life expectancy. Try creating a scatter plot showing the comparison between these two value split by region.

Challenge 3: Solution

- `gdp_percap` as x.
- `life_exp` as y.
- `region` as the legend.



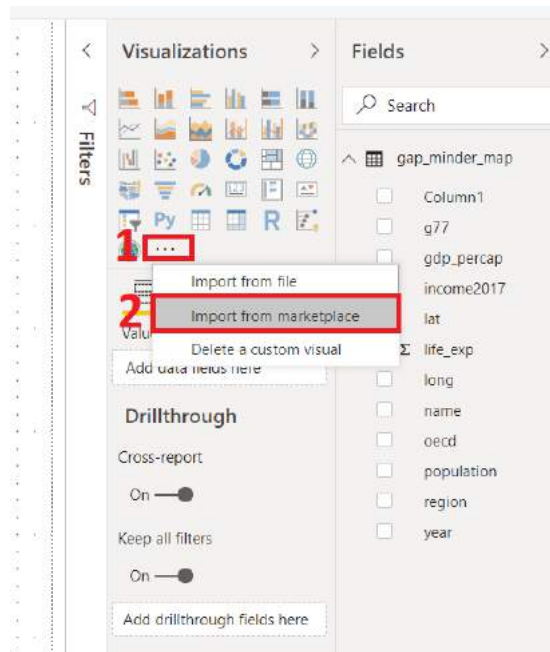
As you can see, this visualization is messy and confusing. We will show ways to make this visualization clearer later on.

3.3 Pages

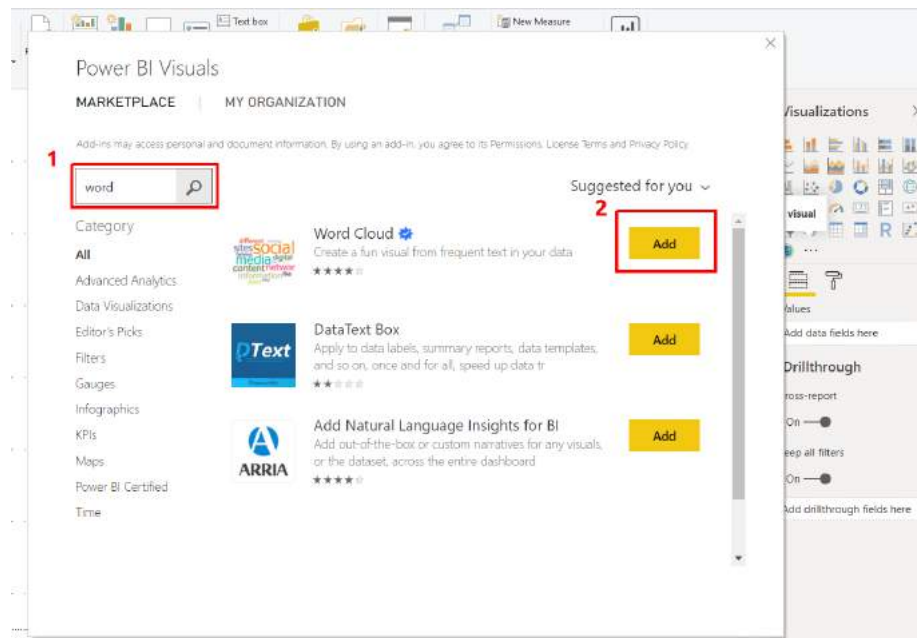
At this stage you will have a lot of visuals but not much room. Power BI has a capability to create multiple pages in a report. When it comes to publishing your report, you can choose to show or hide different pages according to need or use individual visuals for dashboards. Let's start a new page and move the most recent visual you have created on to that page. You can do that by clicking on the visual and then clicking cut on the Home tab of Power BI. Move to the new page and click paste. You can also use CTRL or Command + X and then CTRL or Command + V in the same way as cutting and pasting text in a text editor.

3.4 Explore Marketplace for other visualizations

The visualizations we have been using are part of the default offerings in Power BI. There are more visualisations available in the marketplace. To import a new visualisation template from the marketplace, click on the icon (...), i.e. three horizontal dots in the visualization panel then click on import from the marketplace.



This will pop a new window, where you can search for various types of templates. We are looking for the Word Cloud template. Type Word Cloud in the search box, once you find the template click on the **Add icon**. This will add it to the **Visualization panel**.



Challenge 4: Create a Word Cloud for countries by GDP per capita

- a) Create a plot that compares GDP per capita for each country
- Category: name
 - Values: GDP per capita

Challenge 4: Solution



3.4.1 Map chart

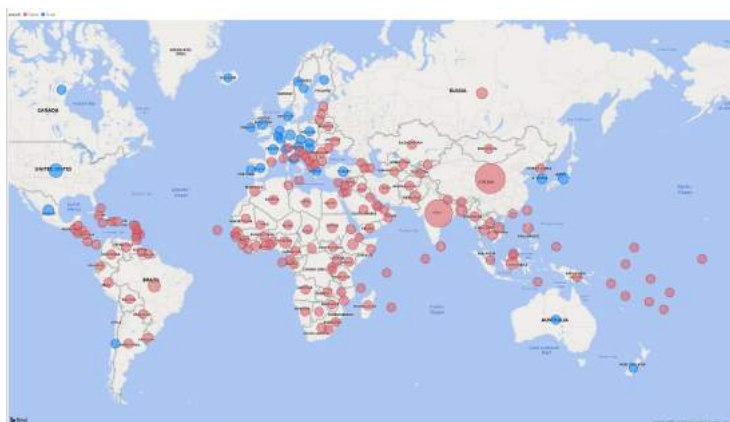
We need Latitude `lat` and Longitude `long` values to create a Map chart. For this, we use the following parameters:

Longitude: `long`

Latitude: `lat`

Size: Population

Legend: `oecd`



This plot tells us if a country was an OECD member by the color of the bubbles. These categories can be seen in the legend at the top of the chart. It also indicates the population size of each country by the size of bubble.

You can also use the `country` field alongside or replacing the longitude and latitude values. Power BI recognises the names of countries however make sure they are spelt correctly or use with longitude and latitude values to ensure accuracy.

3.4.2 Measures and quick measures

At the beginning of the workshop, when we imported the data, we created a calculated column for total GDP using the columns for GDP per capita and population. This is great when we know what columns we will need and if we want to set up our data in a certain way through Power BI when you load the data. However this is not always the case and you will discover you will need certain calculations and measures as you are working with the data. We can do this through the measures function of Power BI.

Select **Quick Measure** on the Home tab. This will give you a variety of different calculations you can use to create measures that can be used in your visualisations. Let's say you want to find the max average life expectancy. Select "Max per category", add `life expectancy` to the base value and add `country` to the category. You will see a new field appear in your **Fields** pane. You will also see a small calculator icon next to the name which indicates this is a measure. Now let's see what this calculates. Create a Matrix visual. This visual is a tabular visual which we can use to show and examine values. Add the new field to the Matrix. A number should now show up that should be the max average life expectancy for a country in the data. We can break this down by adding `region` to the rows to see the max values depending on region. We can then add `country` below region to give a hierarchy that shows the max values for each function.

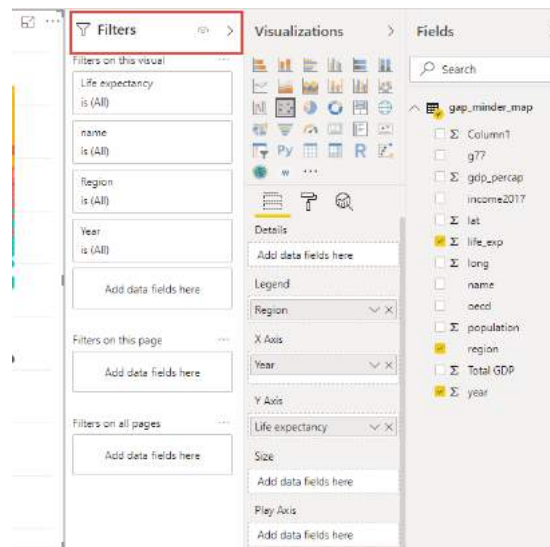
Let's look a bit more at the measure you just created. Click on the field and at the top of the visual, an equation box with an equation should appear. This is the calculation that Power BI is doing based on what you set up in the quick measures prompt. This formula is using the DAX formula language. This is similar to Excel formulas and there are some overlaps so if you are familiar with Excel you will be familiar with a few of the DAX functions. You can edit the formula in your quick measure or write one from scratch using the **New Measure** option on the Home tab. DAX will not be covered in this workshop however you can find the DAX reference guide on the Microsoft Power BI website.

One thing you may notice is that calculating the average for a long period of time, especially when there has been inconsistent data collection like the life expectancy values, might not be meaningful or useful for users looking at your visuals. This is where being able to filter and slice your data is important.

3.5 Filters

Now that we have a set of visuals showing different information, we come to the most important part of data visualisation which is communicating and telling the story of the data. Right now these visualisations, while interesting and potentially useful after some close examination, are difficult to comprehend to the average person or if used in a quick presentation. This is where we will be examining filters.

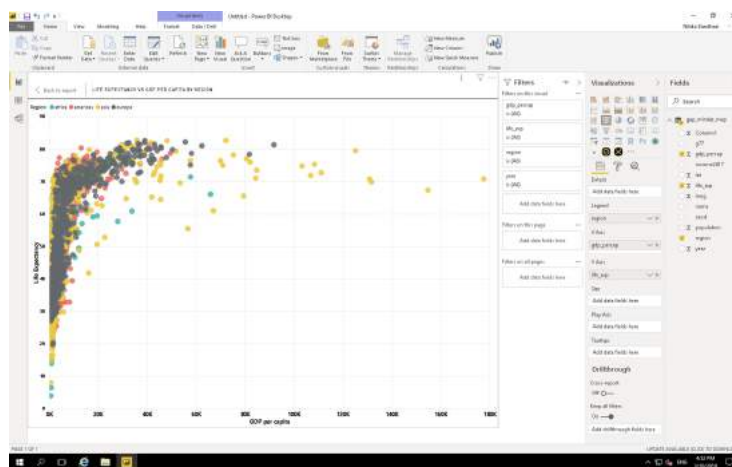
Filters are a powerful way to break down and show specific parts of the data. You can access the **Filters** panel through clicking **Filters**.



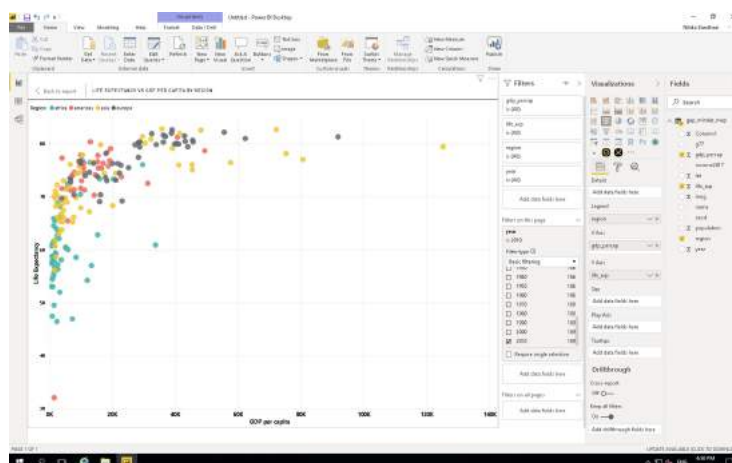
You can filter data for individual visualizations, pages and for the entire report.

Let's have a look at page 2 of our report. This page is rather confusing, we have lots of points on the map for most countries in the world and we have calculations looking at average GDP and life expectancy for 200 years. This is a lot of different information that has been calculated over a diverse timeline.

Select the scatter plot you created for challenge 3. Click on the **Filter panel** which is adjacent to the **Visualization panel**. Drag the **year** field and place it in **Filters on this visual** bar.



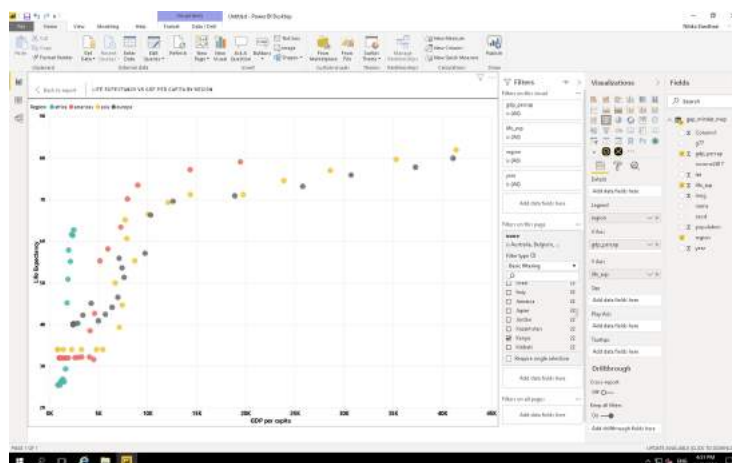
Change the filter type to **Basic Filtering** Now you have all the years as a scrolling list. Select any one year and you can see the data for that particular year. For example, if we want to look at this data from year 2010, select 2010 in the filter panel.



Challenge 5: Filtering

- On the previous plot, create a filter for region.
- Visualize `gdp_percapita` vs `life expectancy` for: Americas

Challenge 5: Solution



This has improved and filtered the data for one visual but what about the rest? This is where the different filter options are important. **Filters on this visual** filters the data only for that visual, the rest of the visuals are unaffected. **Filters on this page** filters the data for all visuals on the page. **Filters on all pages** filters for all pages on the report. Be aware of this option as this will filter data values and options on any new visual you create on any page.

As we would like this page to be focussed on the Americas region in 2010 let's change the above visual filter into a page filter. This will mean that this page will only show data from that region and that year. This will also restrict the end user from looking at other regions or years. What if we would like to give the user some power in exploring the data? This is where slicers come in.

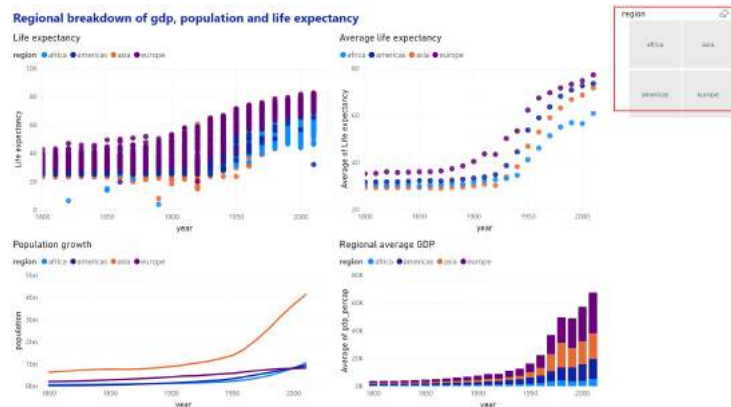
3.5.1 Slicer:

Slicers allow you to put in controls in which the user can explore the data options. NOTE: Slicers are affected by Filters so a Slicer for region on the current page that has a Filter to Americas only will only show the option of Americas.

Let's go back to our first page and create a few Slicers.

In the chart types, locate **Slicer**. Create a new blank Slicer chart. Drag and drop the **region** column in the data field. We have created a slicer and can visualize data for each region separately. Since Power BI is highly interactive, this slicer is applied to all the charts that we have created and will show the results for selection in the slicer. You can chose to apply slicer to specific charts by turning it off for other charts. We will cover this in the next section.

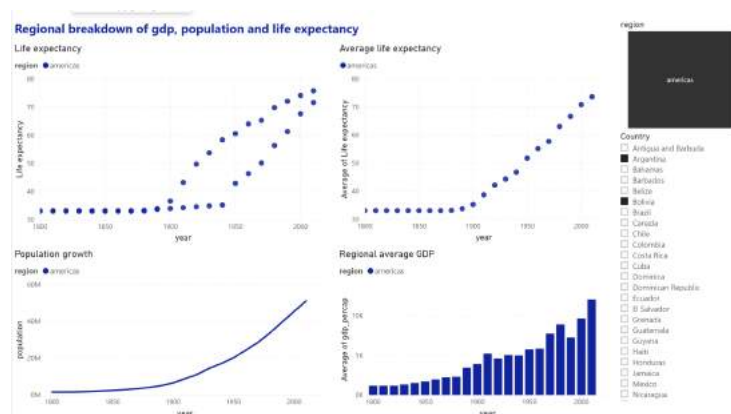
Note: Hold the CTRL button to select more than one option. You can also change the appearance of the slicer the same was as a visual in Format tab. Select General and change orientation to horizontal for clickable buttons instead of checkboxes.



Challenge 6: Slicer

Focussing on the Americas region, create a slicer for countries and examine the differences between Argentina and Bolivia.

Challenge 6: Solution



3.6 Edit interactions

You may have noticed that clicking on a data point in a visualization will affect other visualizations in the same way as a slicer. Click on a segment in your bar chart to see the effect on other visuals.

While this interaction is useful in exploring the connections between the relationships shown in one visual to another, there will be cases where you do not want a visualization to change depending on the slicer or what you have selected on a visual. You can edit how these interactions affect your visualisations through **Edit interactions** on the **Format** tab.

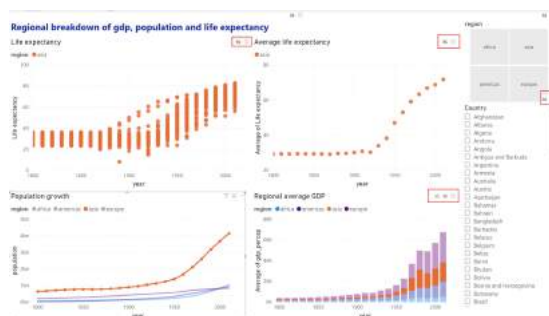
- Select the **population vs year** line chart.

- Select **Asia** on the chart legend.
- Select the **Format** menu and select **Edit interactions**.

Several icons will pop up near the other visualizations. These control how the visual will look depending on your interaction with the selected visual.

Selecting **None** will mean that the related visual will not change depending on interactions with the selected visual. Change one of the scatter plots to **None** and select a different data point on the line chart to see the effect. Select **Asia** on the line chart again. Currently the bar chart highlights the relevant section. Select the filter icon to change and select another data point on the line chart to see the effect.

Note: existing elements on the page may interfere with the **Edit interactions** icons. Resize and move your visuals to avoid this.



This also affects slicers, selecting the slicer will show the same **Edit interactions** icons. You may not want the average life expectancy by region visual to change depending on which countries are chosen. Select the country slicer and select **None** on the visual to stop that visual changing due to the options on the slicer.

Every visual interacts with all other visuals individually. Check the **Edit interactions** to make sure all your interactions work as expected.

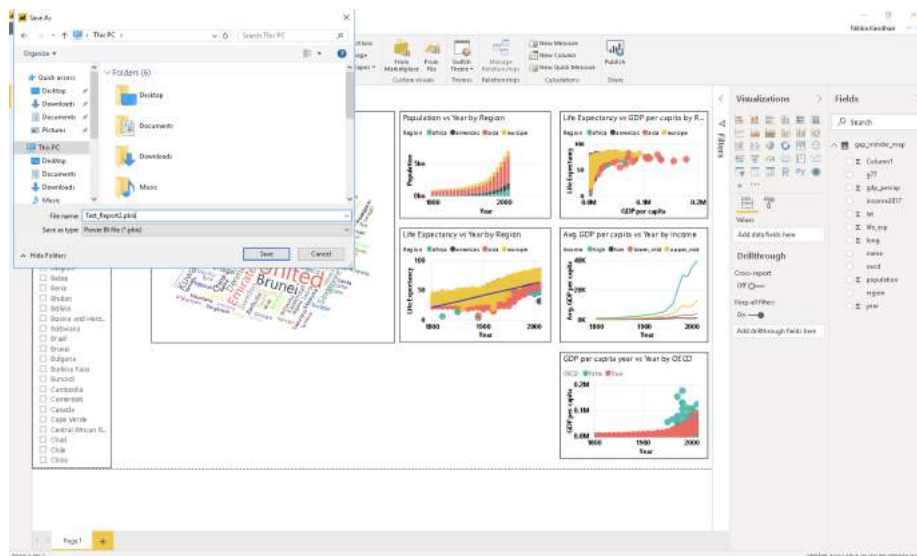
Chapter 4

Saving and Exporting

Now that we have visualised the data and created various types of plots we need to save our visuals.

4.1 Saving as pbix files

Go to **File > Save**, then choose an appropriate file name and the folder to save it. Click on **Save**.



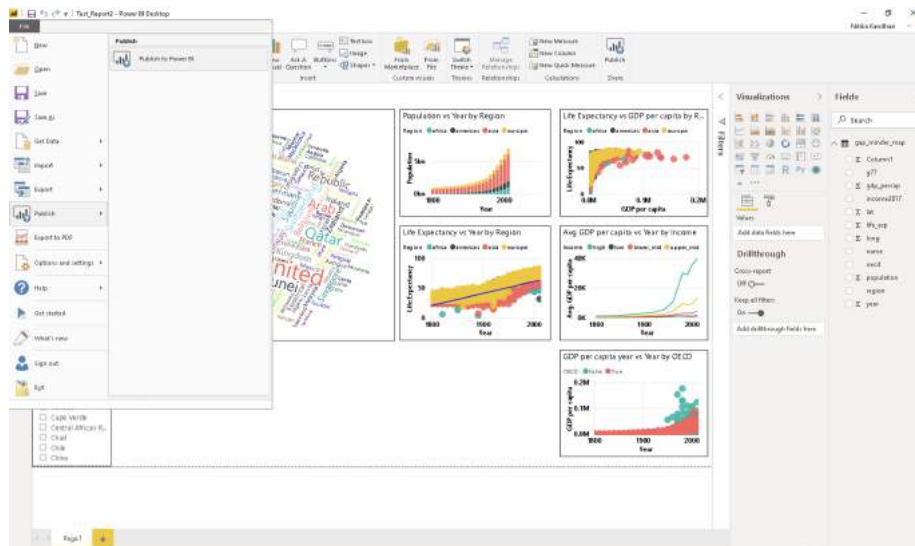
The file extension for Power BI files is “.pbix”. The .pbix files are highly compressed file types that contains all the graphics along with the actual data.

Note: If you share your .pbix file with others, your actual data gets shared as well.

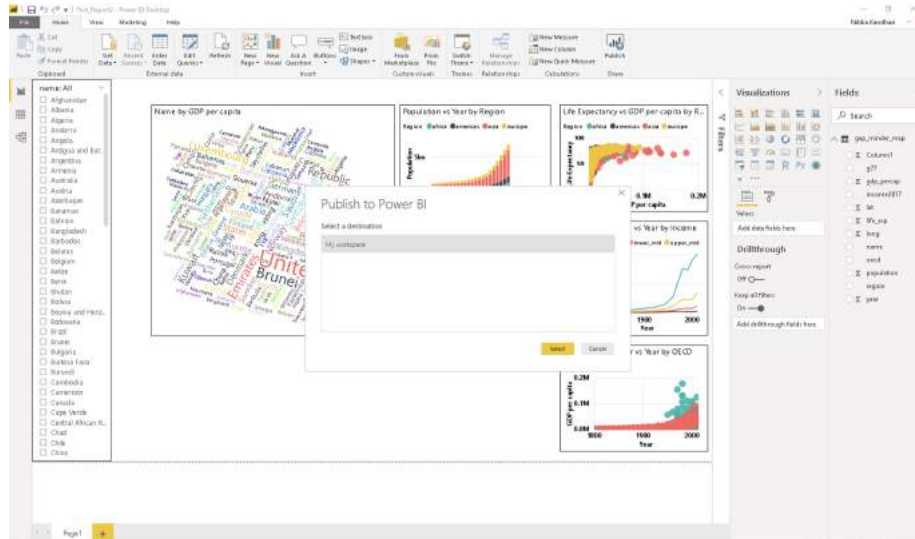
4.2 Exporting and publishing report

Now we want to create a report of our analysis and share it with others. We use Power BI to publish the report on the Power BI server.

To do this, go to **File > Publish > Publish to Power BI**.

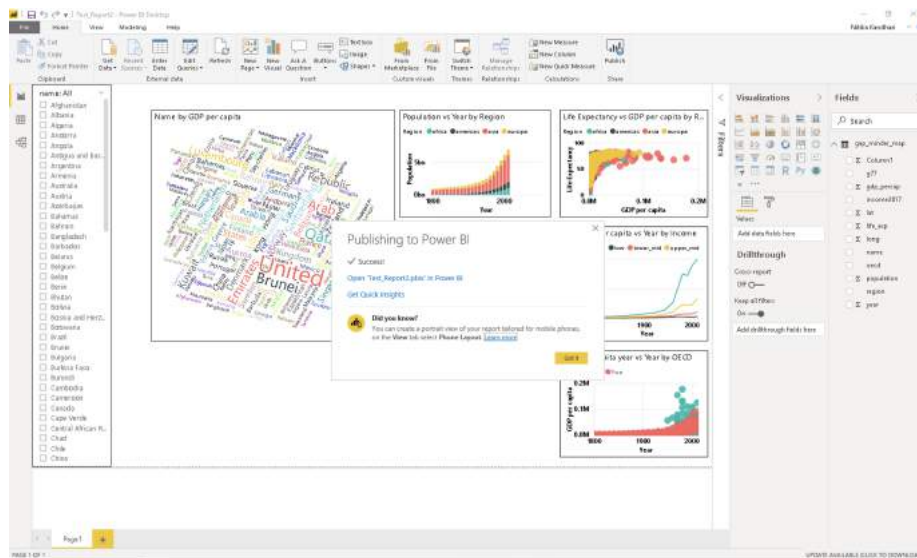


After clicking on **Publish to Power BI**, you will be prompted to select a destination on the Power BI server where you want to publish your report. The default setting is **My workspace**. You can create different workspaces dedicated to different projects on the server. Let's use the default option, and click **Select**.

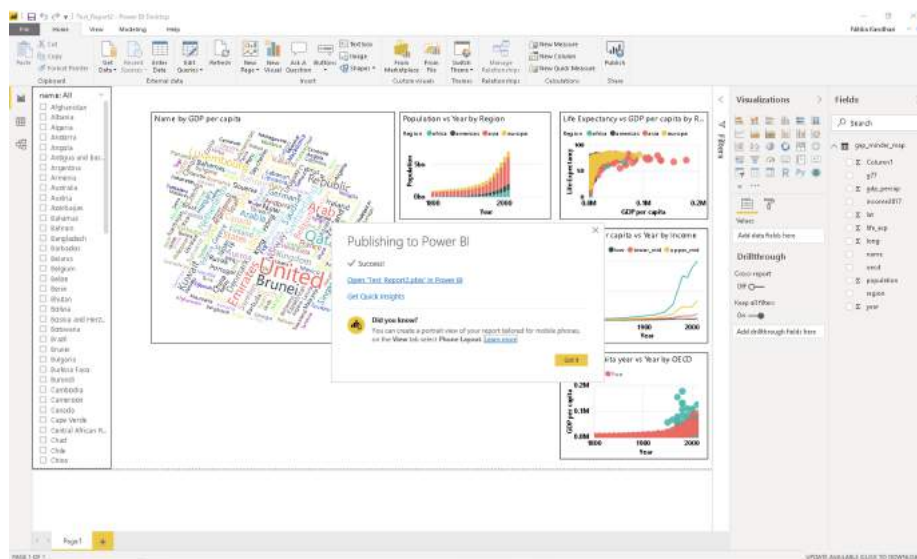


It takes about 2-3 mins to finalise the report and you will see the following success message on your screen.

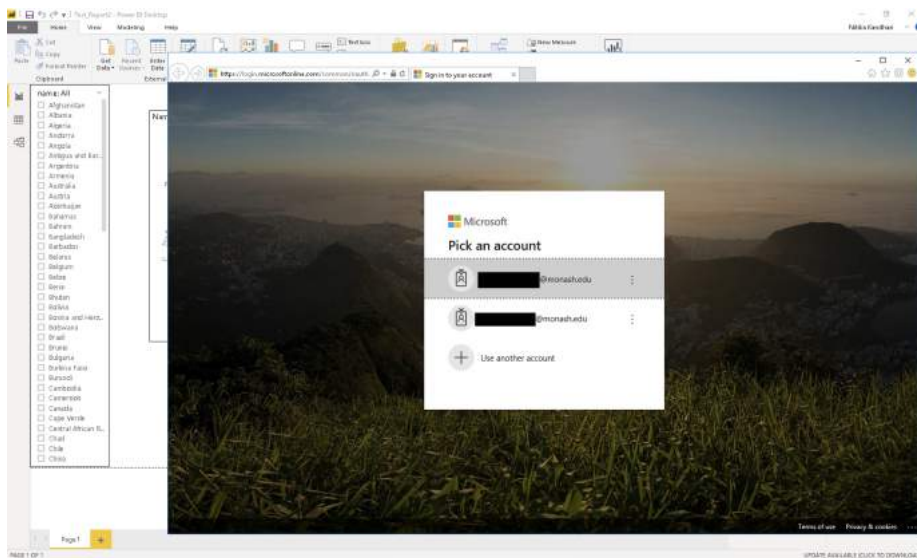
CHAPTER 4. SAVING AND EXPORTING



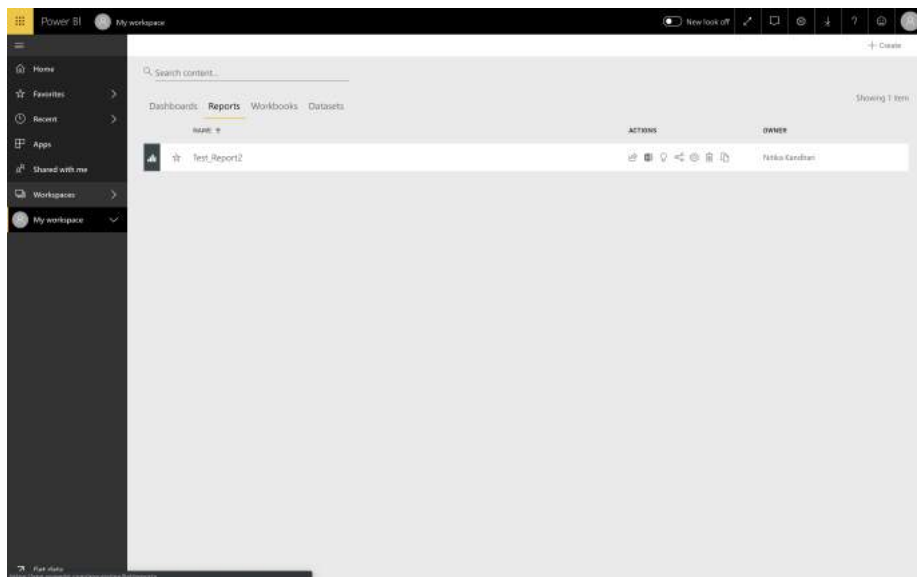
Congratulations! The report has been published to your workspace on Power BI. To access it, click on **Open <YOUR_FILENAME.pbix>** in Power BI.



After you click on the link, you will be taken to Power BI's login through Microsoft. You need to fill in your login credentials.

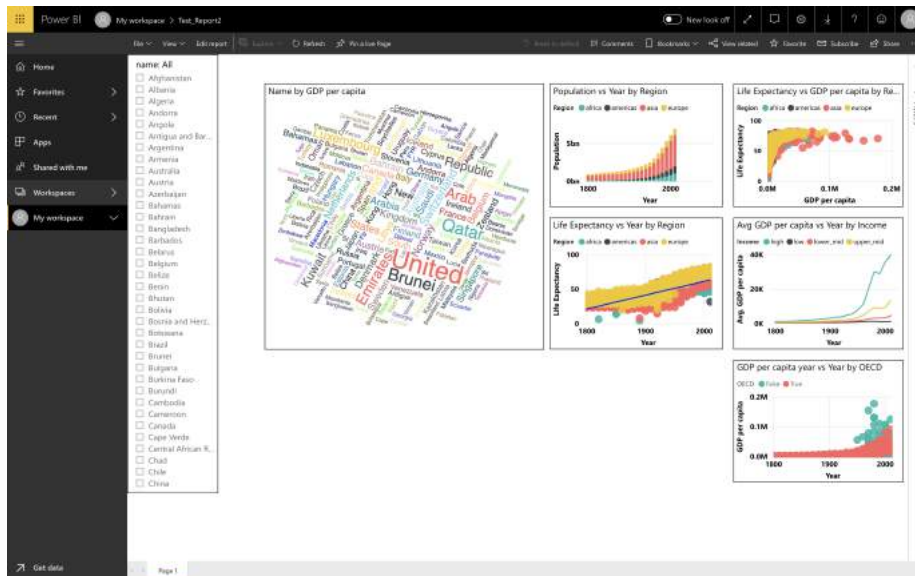


Once you sign in, you will see the Power BI dashboard. Scroll through the tabs on the top - Dashboards, Reports, Workbooks and Datasets. We are interested in the **Reports** tab because that is where you can find your published report. The different types of actions can be seen under the **Actions** tab. These actions include - share, analyze in Excel, quick insights etc. Hover over the different symbols to find out what they do and select the appropriate action.

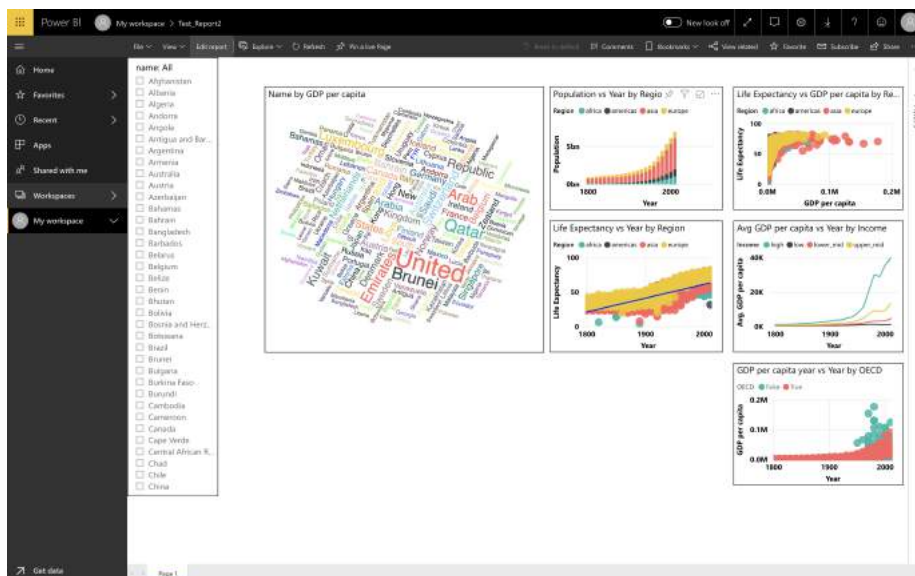


We want to take a look at our report and share it with others. Let's click on the report itself. You will see a similar screen showing your report:

CHAPTER 4. SAVING AND EXPORTING

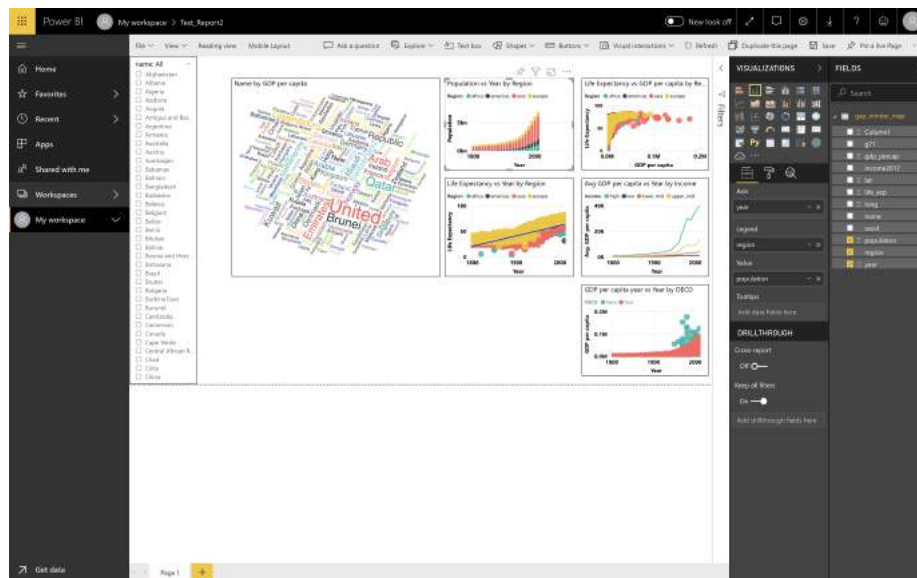


If you click on **Edit** on the top, you will be able to edit mode and perform nearly everything you can in the Power BI desktop.



You can see all the Visualization options and fields in the Power BI server. Click on the **Share** button in the top left corner. This allows you to share your report with others.

CHAPTER 4. SAVING AND EXPORTING



Chapter 5

Next steps

5.1 Deepen your understanding

Microsoft has in-depth guides and tutorials on their Power BI Guided Learning¹ website.

You can visit the Microsoft Power BI Blog² for the latest updates on performance improvements, more visualization options and so on. There is a community of forums and blogs for you to discuss with others.

Join the Power BI Network Forum on Workplace³ for the latest discussions around Power BI at Monash University.

5.2 Join the community

Join the Data Fluency community at Monash⁴.

- Mailing list for workshop and event announcements.
- Slack for discussion.
- Monthly seminars on Data Science topics.
- Drop-in sessions on Friday afternoon.

The Carpentries⁵ run intensive two day workshops on scientific computing and data science topics worldwide. The style of this present workshop is very much based on theirs.

¹https://docs.microsoft.com/en-us/power-bi/guided-learning/?WT.mc_id=PBIService_GettingStarted

²<https://powerbi.microsoft.com/en-us/blog/>

³<https://monash.workplace.com/groups/730041584039981/>

⁴<https://www.monash.edu/data-fluency>

⁵<https://carpentries.org/>