

## Data Visualisation with Power BI

87% Complete  
Last activity on December 31, 2020

### Introduction to Power BI

<a href="#">Course Introduction</a>	<a href="#">1 Topic</a>
<a href="#">Introduction to Power BI</a>	<a href="#">1 Topic</a>
<a href="#">Visualisations in Power BI</a>	<a href="#">2 Topics</a>
<a href="#">Dashboards in Power BI</a>	<a href="#">3 Topics</a>

### Correlation

<a href="#">Intro to Correlation</a>	<a href="#">1 Topic</a>
<a href="#">Correlation in Excel</a>	<a href="#">2 Topics</a>
<a href="#">Correlation in Power BI</a>	<a href="#">3 Topics</a>
<a href="#">Feedback point – how's the course going?</a>	

### Advanced Visualisations

<a href="#">R and Python Visualisations</a>	<a href="#">1 Topic</a>
<a href="#">Text Analysis with Word Clouds</a>	<a href="#">5 Topics</a>

### Assignment

<a href="#">Formative Assessment – Craft Beer</a>	<input type="radio"/>
<a href="#">Formative Assessment – Craft Beer</a>	
<a href="#">Summative Assessment – World Happiness</a>	<input type="radio"/>
<a href="#">Course Feedback</a>	<input type="radio"/>

### Quizzes

<a href="#">Data Visualisation with Power BI   Final Quiz</a>
---

## Formative Assessment – Craft Beer



Jimmy Wynen September 22, 2020

### Results

23 of 25 questions answered correctly

YOUR TIME:  
01:54:07

You have reached 23 of 25 point(s), (92%)

[Restart Quiz](#)

[View questions](#)

[Click Here to Continue](#)

Welcome to this formative, or practice, assessment.

**Aim:** The purpose of this lesson is to help you prepare for the summative assessment in a similar environment.

### Method:

In this formative assessment, you'll put into practice the skills you've learnt within the Data Visualisation with Power BI course.

During this assessment, you'll be presented with various questions or problems that you should attempt on your own. The answer will then be presented on the next page, with a step-by-step walkthrough (with screenshots or gifs).

**Materials:** You'll require the two datasets below:

- dataset\_beers
- dataset\_breweries





Darius S



\_gary.smith



\_Aaron Harmer



\_AlanAndres



Aaron

[Show more](#)

These are also found in the 'Materials' tab.

**Scenario:**

You are a data analyst at a craft beer company. The company wants to check trends in craft beer & analyse what styles are popular in order to decide what products to produce. They require this data in the form of an interactive dashboard, and it's up to you to deliver it!

**Terminology:**

Using this dataset and scenario you'll be introduced to the wonderful world of craft beer. If you're a craft beer fan (like myself) you'll likely be already familiar with the below terms.

**ABV:** Alcohol by volume. What alcoholic percentage is the beer? Beers *usually* fall between the 4-6% mark, with some styles (such as IPA) being (on average) higher in ABV.

**IBU:** International Bittering Units. How bitter is the beer? Styles such as lagers or wheat beers will generally be low in IBU, whereas styles such as IPA (India Pale Ale) will be more bitter.

**Style:** Beers come in a variety of different styles! They range from simple light-coloured lagers to dark-coloured stouts and porters.

**Ounces:** As this is an American dataset their beer is measured in ounces. If you like, you can change this with a calculated column later on.

More details on beer measurement can be found at the link below:

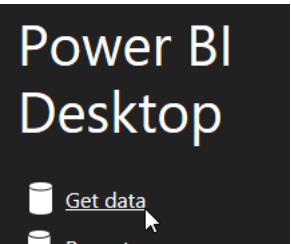
[https://en.wikipedia.org/wiki/Beer\\_measurement](https://en.wikipedia.org/wiki/Beer_measurement)

*Take a quick look at the assessment workbook and scenario provided, then click into the quiz to begin!*

☒ Let's get started!

Correct

Firstly let's load our data into Power BI. Open Power BI Desktop and click 'Get Data'.



Select 'Text/csv' then 'Connect'.

## Get Data

All

File

Excel

Text/CSV

*Note: If you move, delete, rename or otherwise change your data it will cause issues to your Power BI file. Avoid this at all costs – I recommend having a folder that you don't touch with these datasets within.*

Load the first dataset, **dataset\_beers.csv**, into Power BI.

Take a look at the column headings and think about the data we have per row & what metrics we might be able to use to analyse our data.

At this stage, it would be best practice to explore the data, it's limitations and creates a data dictionary. If you'd like the practice, **create a data dictionary for this dataset** within Excel, or a Word or Google doc, & I'll take you through the one I've created on the next page.

dataset\_beers.csv

File Origin: 1252: Western European (Windows)
Delimiter: Comma
Data Type Detection: Based on first 200 rows

	abv	ibu	id	name	style	brewery_id	ounces
0	0.05	null	1436	Pub Beer	American Pale Lager	408	12
1	0.066	null	2265	Devil's Cup	American Pale Ale (APA)	177	12
2	0.071	null	2264	Rise of the Phoenix	American IPA	177	12
3	0.09	null	2263	Sinister	American Double / Imperial IPA	177	12
4	0.075	null	2262	Sex and Candy	American IPA	177	12
5	0.077	null	2261	Black Exodus	Oatmeal Stout	177	12
6	0.045	null	2260	Lake Street Express	American Pale Ale (APA)	177	12
7	0.065	null	2259	Foreman	American Porter	177	12
8	0.055	null	2258	Jade	American Pale Ale (APA)	177	12
9	0.086	null	2131	Cone Crusher	American Double / Imperial IPA	177	12
10	0.072	null	2099	Sophomoric Saison	Saison / Farmhouse Ale	177	12
11	0.073	null	2098	Regional Ring Of Fire	Saison / Farmhouse Ale	177	12
12	0.069	null	2097	Garce SeilAB	Saison / Farmhouse Ale	177	12
13	0.085	null	1980	Troll Destroyer	Belgian IPA	177	12
14	0.061	60	1979	Bitter Blitch	American Pale Ale (APA)	177	12
15	0.06	null	2318	Ginija Ninja	Cider	154	12
16	0.06	null	2170	Cherried Away	Cider	154	12
17	0.06	null	2169	Rhubarbarian	Cider	154	12
18	0.06	null	1502	BrightCider	Cider	154	12
19	0.082	null	1593	He Said Baltic-Style Porter	Baltic Porter	368	12

The data in the preview has been truncated due to size limits.

Load
Transform Data
Cancel

Next!

Correct

Each row is data on a single beer. Your data dictionary should look similar to the below:

Field Name	Type	Description
------------	------	-------------

no column heading / column1	Number	Index number
abv	Number (between 0-1)	ABV (Alcohol by Volume) of the beer
ibu	Number	IBU (International Bittering Units) of the beer
ID	Number	Unique ID for the beer/row of data
Name	Text	Name of the beer
Style	Text	Style of the beer (101 distinct)
brewery_id	Number	ID of the brewery that brewed the beer – links to dataset_breweries.csv
ounces	Number	How many ounces is a can or bottle of the beer

In this dataset, which columns might we be able to measure? (by looking at descriptive statistics like count, average, % of total, etc). *Select all that apply*

☒ ABV

☒ IBU

☐ ID

☐ Name

☐ Style

☐ brewery\_id

☒ Ounces

Incorrect

Great, so we could analyse most of these columns in different ways.

What are some of the questions for targeted analysis? (select all that apply).

*Note: this is not covering off all of the questions, there are indeed many more we could look at – this is just to get us started*

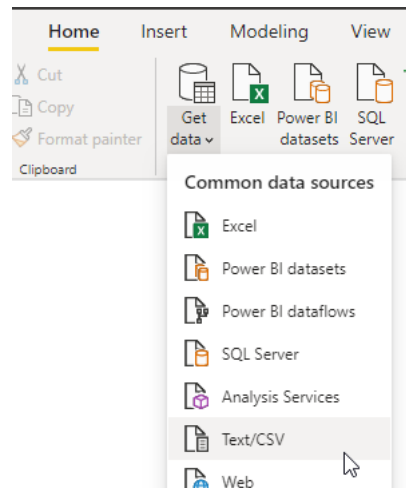
- ☐ What is the sum of ABV?
- ☒ What is the average ABV?
- ☐ What is the average brewery\_id?
- ☒ What is the average IBU per style?
- ☒ Is there a correlation between ABV and IBU?
- ☒ What is the most popular style (Count or % of Total)

**Correct**

Awesome, so generally we'll be looking at measures of central tendency for ABV & IBU, per style, and how we can visualise this in an interactive dashboard.

Next load in dataset\_breweries.csv

(on the home tab > Get data > Text/CSV)



File Origin

Delimiter

Data Type Detection

1252: Western European (Windows)

Comma

Based on first 200 rows

	name	city	state
0	NorthGate Brewing	Minneapolis	MN
1	Against the Grain Brewery	Louisville	KY
2	Jack's Abby Craft Lagers	Framingham	MA
3	Mike Hess Brewing Company	San Diego	CA
4	Fort Point Beer Company	San Francisco	CA
5	COAST Brewing Company	Charleston	SC
6	Great Divide Brewing Company	Denver	CO
7	Tapistry Brewing	Bridgman	MI
8	Big Lake Brewing	Holland	MI
9	The Mitten Brewing Company	Grand Rapids	MI
10	Brewery Vivant	Grand Rapids	MI
11	Petoskey Brewing	Petoskey	MI
12	Blackrocks Brewery	Marquette	MI
13	Perrin Brewing Company	Comstock Park	MI
14	Witch's Hat Brewing Company	South Lyon	MI
15	Founders Brewing Company	Grand Rapids	MI
16	Flat 12 Bierwerks	Indianapolis	IN
17	Tin Man Brewing Company	Evansville	IN
18	Black Acre Brewing Co.	Indianapolis	IN
19	Brew Link Brewing	Plainfield	IN

The data in the preview has been truncated due to size limits.

Load

Transform Data

Cancel

We can see that this dataset shows the locations of breweries within the US. What visualisations within Power BI represent an efficient way to visualise our measurable values against the locations in our dataset?

☐ Doughnut

☐ Funnel

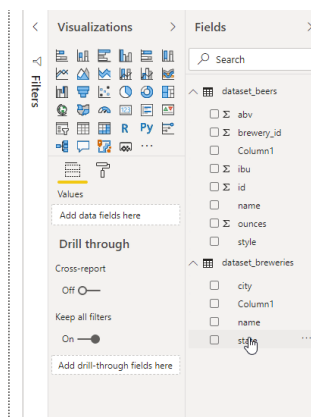
☐ Waterfall

☒ Map or Filled Map

Correct

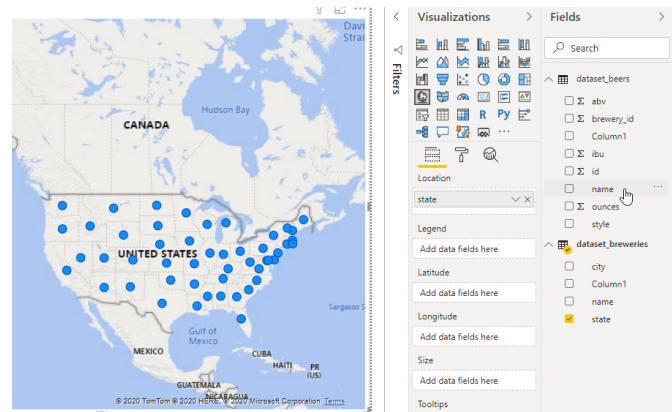
Looking at our breweries dataset, which contains location data, we could visualise this on a map or filled map visualisation.

Once your data is loaded in, click and drag the "State" column from your fields list into the reports pane. This will create a Map visualisation.

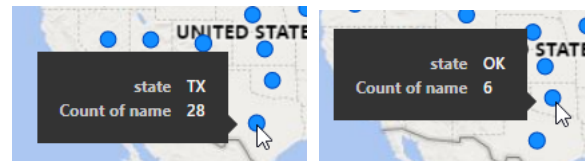


While this shows us the states that exist in the dataset, it's not very useful at the moment (turns out all the states brew beer!). Let's update our visualisation so the size of each bubble represents how many beers are brewed in each state.

With your map visualisation highlighted, Click and drag "name" from the fields list into "Size".



Hang on, something is not quite right. The size of our bubbles on the chart hasn't changed at all, but if we hover over Texas (TX) for example, we can see that there's a count of 28 – which should have a larger bubble than Oklahoma (OK) state (just above).

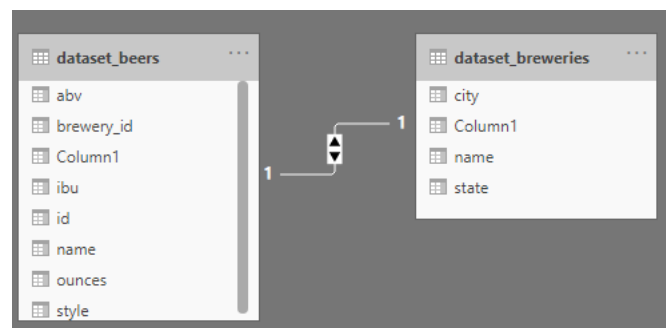


Let's take a look at our data model. Click into the **model** tab

☒ I've clicked into the model tab

Correct

In our **model**, we can see a 1 to 1 relationship has been automatically created:



Hover over the relationship and what do you see. What could be the issue here?

☒ We have a 1:1 relationship which is incorrect, this should be many to many

☐ There should be multiple relationships

☐ The relationship in dataset\_beers is based on

"Column1" which is just an index column

Incorrect

What should the relationship be?

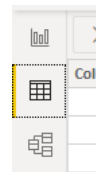
#### Sort elements

Type of relationship	Many to 1
Dataset_Beers	brewery_id
Dataset_breweries	Column1

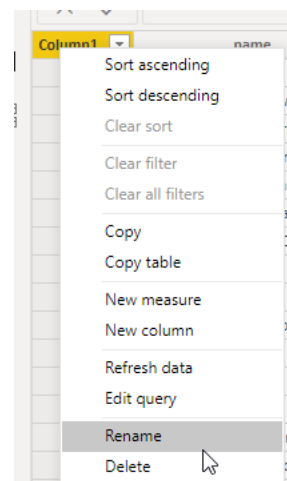
Correct

Within dataset\_breweries the column1 should be renamed to "brewery\_id" for consistency. Let's update that first.

Click into the data tab on the left-hand side of Power BI.



Click into dataset\_breweries on the right-hand side, then right-click on column1, and select Rename.



Directly type in "brewery\_id" and press enter.

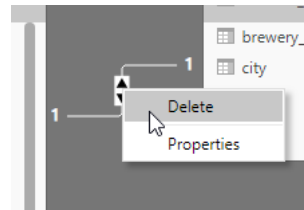




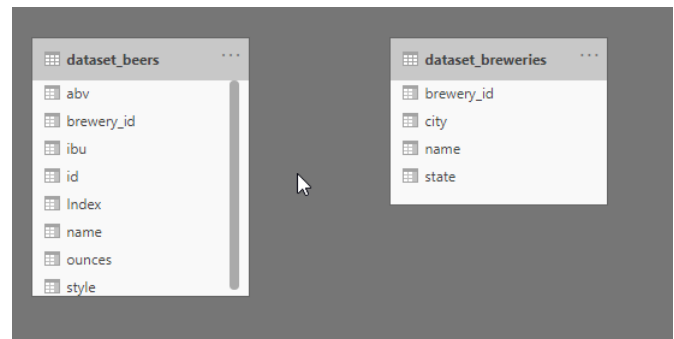
brewery_id	name	city	state
0	NorthGate Brewing	Minneapolis	MN
1	Against the Grain Brewery	Louisville	KY
2	Jack's Abby Craft Lagers	Framingham	MA
3	Mike Hess Brewing Company	San Diego	CA
4	East Point Beer Company	San Francisco	CA

While we're here, rename the other "column1" within the dataset\_beers to "index".

Once that's done, let's jump back into the model and let's update the relationship. First, delete the current relationship by right-clicking and select "delete".



Then, create a new relationship by clicking and dragging brewery\_id from dataset\_beers onto brewery\_id in dataset\_breweries.



This has created a many to 1 relationship.

Each beer is created by a single brewery – and each brewery creates many beers. Hence, many to 1.

You can hover over the arrow to highlight which fields are linked, and you can right-click > properties to see further information on the relationship.

Check that 'Make this relationship active' is ticked.



## Edit relationship

Select tables and columns that are related.

dataset\_beers

Index	abv	ibu	id	name	style	brewery_id	ounces
2	0.071	null	2264	Rise of the Phoenix	American IPA	177	12
4	0.075	null	2262	Sex and Candy	American IPA	177	12
47	0.065	null	2286	Divided Sky	American IPA	60	12

dataset\_breweries

brewery_id	name	city	state
0	NorthGate Brewing	Minneapolis	MN
1	Against the Grain Brewery	Louisville	KY
2	Jack's Abby Craft Lagers	Framingham	MA

Cardinality
Cross filter direction

Many to one (\*:1)
Single

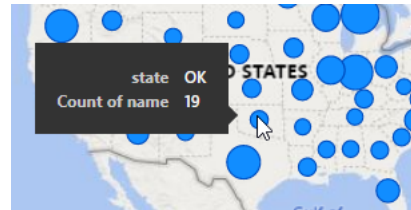
☒ Make this relationship active
☐ Apply security filter in both directions
☐ Assume referential integrity

OK
Cancel

Next!

Correct

Move back to the Reports tab, and we can now see that the size of the bubbles on our maps chart is correctly showing – where states that have brewed more beers have a larger bubble.



Instead of showing “Count of name”, however, let’s rename this to “Number of beers brewed”. You can do this by simply clicking on the chevron (dropdown arrow) and selecting “Rename”, then typing in directly.

Add data fields here

Size

Count of name

Tooltips

Add data fields here

Drill through

Using this visualisation, where the size of the bubble indicates the number of beers brewed in that state, which state has the most number of beers brewed?

AK

CO

☐ NY

☐ TX

Correct

Let's take a look at the most popular style of beer.

On the same report page, create a bar chart: **Count of Name by Style.**

1. Create a Bar chart by clicking on the visualisation
2. Click and drag Name into Values, this should automatically aggregate to Count of Name.
3. Click and drag Style into Axis

Which style has the most beers brewed?

☐ American Pale Ale (APA)

☒ American IPA

☐ Cider

☐ American Brown Ale

Correct

Use the same bar graph to answer the following question:

How many beers were produced of the 3rd most popular style?

*Hint: You can hover over the bar for the tooltip*

☐ 105

☐ 108

☒ 133

☐ 245

Correct

Now that you've created this bar chart, you can use this as a slicer to interact with your map visualisation on the same page.



Scroll to the bottom of your bar chart, where “Wheat Beer” is at the bottom of the chart, & click on it. Which state produces the only Wheat Beer in this dataset?

☒ CA

☐ CO

☐ OR

☐ WA

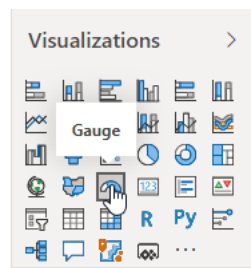
**Correct**

Let’s put in a couple of gauge visualisations to look at ABV and IBU, and see which styles are above or below the total average.

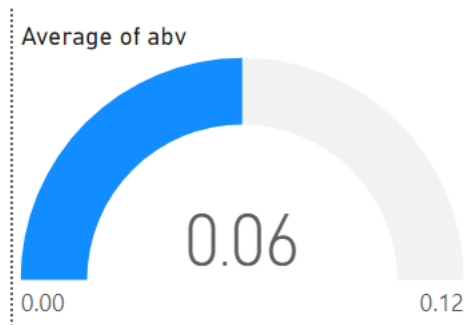
The gauge visualisation, if you haven’t used it before, is a good way to visualise:

1. Tracking progress towards a milestone (such as a sales target)
2. Visualising above or below a certain threshold (such as the total average of a particular metric when filters are applied).

The Gauge is found here:



Click and drag ABV into the ‘Value’ section, then change it from the default SUM to average.



Great, but wouldn’t it be better if it displayed as a percentage? Let’s update that.

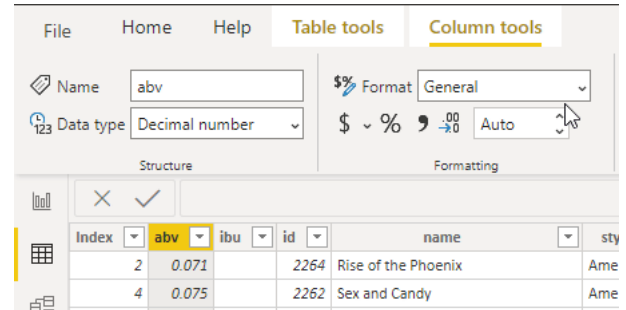


☒ I'm ready to update!

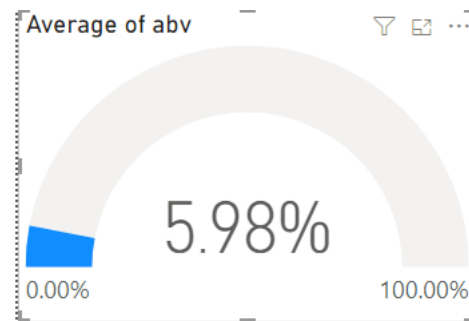
Correct

Go into the Data tab, then click on the ABV field on the right-hand side. ABV will now be highlighted in yellow in your table.

On the Column tools tab click on the format dropdown and change it to 'Percentage' – refer to the gif below.



With this done we can click back into the report tab and see our gauge updated like so:



Awesome, but the maximum ABV isn't going to be anywhere close to 100% – so let's optimise our gauge.

☒ Let's optimise!

Correct

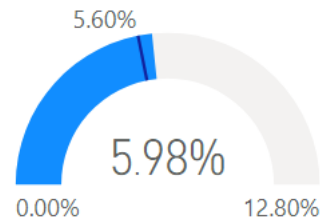
One way we could display this is by dragging ABV into the Maximum value section and changing it from the default SUM to a MAX. This will then show the **average ABV** per beer as the blue data (& the number in the middle) and the **maximum value** as the maximum value on the gauge.

Click on some different styles and see how the gauge updates – you can see how some averages are much closer to the max value.

Let's add in a target value – again click and drag ABV, this time into the Target Value section, then click on the dropdown and change to **Median**.

Your gauge should look similar:

Average of abv, Max of abv and Median of abv



Now that we have both the median and the mean (average) plotted, this is a handy way to check the skew of the dataset. If there's a large difference between the median and the mean, this could mean a skewed dataset.

☒ Next!

Correct

We've created one way of using a gauge, but for our final dashboard, we'd like to use the gauges to see when different styles are above or below the total average.

To achieve this, we'll look at another way to set the maximum and target values. Delete this gauge visualisation, and create another to show average ABV – but this time, we'll set the target line to the total ABV (across all styles), that way we can easily determine higher or lower ABV than the average when we slice our data.

Create a new gauge visualisation. Click and drag ABV into Values, then set this to Average of ABV.

Next, click into the format tab and click on the dropdown 'Gauge Axis'. Set the target to 0.0598 (which is the average ABV) and the Maximum to 0.128 (which is the maximum ABV).

For the style, 'American Pale Ale (APA)', is the average ABV higher or lower than the total average? (use the gauge to determine).

☒ Lower



☐ Higher

Correct

Create another gauge visualisation, this time for IBU. Complete the same steps as before, with the average value as the target and the maximum value as the maximum.

For the style, 'American Pale Ale (APA)', is the average IBU higher or lower than the total average? (use the gauge to determine).

☐ Lower

☒ Higher

Correct

Rename this report tab to 'Dashboard', and create a new report tab by clicking on the (+) button.

*Note: to rename a report tab, simply double click on the current name.*

On this page we'll take a look at another question:

◦ **Is there a correlation between ABV and IBU?**

Which visualisation can we use to plot the relationship between ABV and IBU?

☐ Waterfall

☐ Bar Chart

☐ Pie Chart

☒ Scatter Plot

Correct

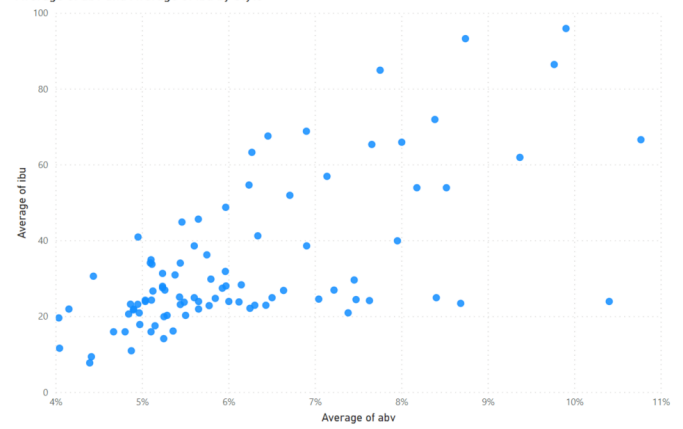
Great, we'll use a scatter plot to visualise the relationship between the two variables, ABV and IBU.

Click and drag ABV into the X Axis, then use the dropdown to change it to Average. Do the same for IBU on the Y Axis, then bring Style into the Details section.

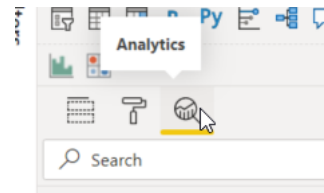
You should have something like this:



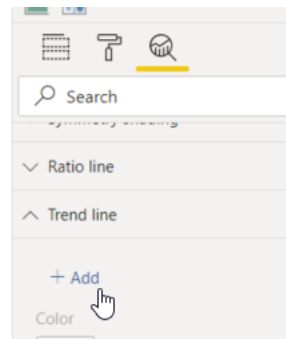
Average of abv and Average of ibu by style



Let's add a trendline to better visualise the relationship. With the scatter plot selected, click into the Analytics tab:



Click into the 'Trend Line' dropdown, then click '+ Add'.



Great, with our trendline what can we determine about the relationship between ABV and IBU?

☒ As ABV increases, IBU increases

☐ As ABV increases, IBU decreases

Correct



There's other lines we can include too, like Min, Max and Median values, or an X or Y axis constant line. Let's look at adding a couple of percentile lines to see the beer styles that are highest in both ABV and IBU.

Scroll down to 'Percentile line' and click '+ Add'. Select the measure 'Average of IBU' and change the Percentile to 95%. Add another line for 'Average of ABV', with the same 95% percentile.

Which beer styles are above 95% in both ABV and IBU?  
(Average) *Select all that apply*

- ☒ American Barleywine
- ☒ Russian Imperial stout
- ☐ American Double / Imperial IPA
- ☐ English Barleywine
- ☐ American Double / Imperial Stout

**Correct**

There's a large number of styles here, so let's try and group some styles together. We'll group beer styles from a particular country together – for example, let's group American IPA, American Pale Ale, etc together under the banner 'American'.

Jump into the data tab, and in the dataset\_beers data create a new column and copy/paste the formula below. This is an IF statement that will assist with grouping.

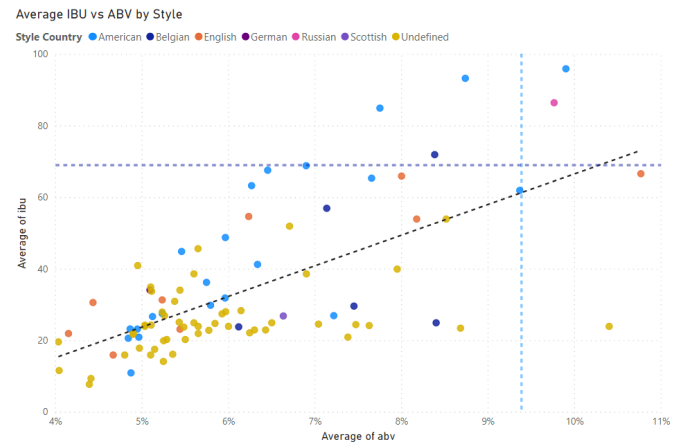
StyleCountry =  
IF(CONTAINSSTRING(dataset\_beers[style],"American"),"American",IF(CONTAIN

The formula above is pretty long, but essentially it's a big IF statement – IF the style includes the word "American", then it'll be American. If it includes the word "English", then it's English. Etc etc – if it doesn't include any of the specified words, then it will be Undefined.

With that done, click back into the reports tab and bring your new 'StyleCountry' field into the Legend section of your scatter plot.

This will group the styles into different colours based on this new field. Your visualisation should look a little like the below. Note I've updated the wording of the title and the legend.



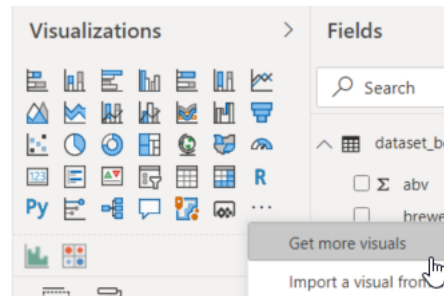


☐ Next!

**Correct**

Next, let's plot the correlation coefficient of ABV vs IBU using an additional visualisation – we used this before in the Correlation lesson series.

If you don't already have it, you can click on the '...' in the bottom right-hand corner of the visualisations > Get more visuals.



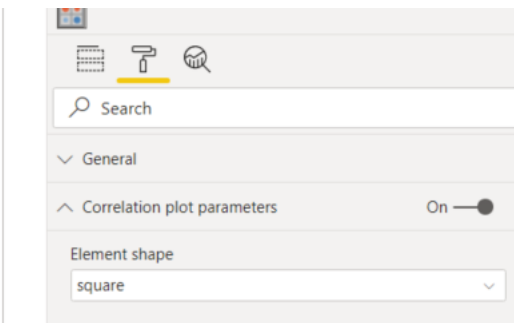
Then search for 'correlation', choose the correlation plot and 'Add'.

You'll then have the correlation plot visualisation imported. Select it to add it to our report.

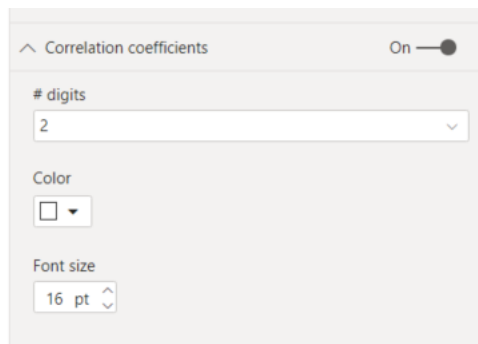
Next, click and drag both ABV and IBU to the 'Values' section. Based on this chart, we can see that ABV has a strong correlation with IBU (as we suspected!).

We can optimise this chart a little better though, so let's jump into the format tab. First, under 'Correlation plot parameters', change the Element shape to Square.





Next, turn on 'Correlation coefficients', change the # digits to 2, change the color to white, and bring the font size up to 16pt.



Your visualisation should now look similar to this:



☒ Awesome!

☐ Eh

Correct

Great, now we've plotted the correlation between ABV and IBU – and it's fairly strong as the trendline indicated.

Lastly, let's jump back into the dashboard where you have your map, bar chart and gauges.

Let's update the style bar chart with our StyleCountry so we can drill down.



Click and drag StyleCountry into the Axis, above Style.

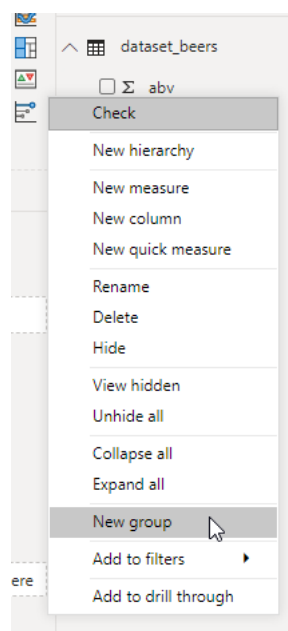
This will update the bar chart and you can now drill down using the downwards arrows or by right-clicking on any of the bars.

☒ It's looking good!

**Correct**

As our last Visualisation, let's create a histogram based on ABV. We'll achieve this by creating a Group from our ABV field.

Click on the (...) on ABV, then choose 'New Group'.



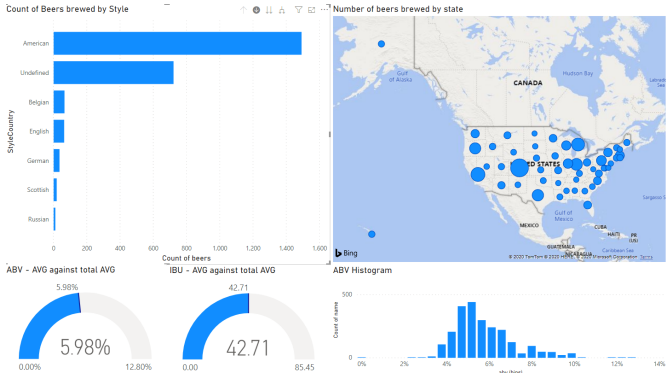
Leave the defaults for now, then click OK. You'll now have a new field – abv (bins) which we can use.

Create a column chart with abv (bins) on the axis and Count of name as the Values. Change the title to: ABV Histogram.

If you like, you can edit the group to resize your histogram's classes. Play around with your dashboard first & choose some different styles to see the spread of values.

Do some final formatting by resizing some of your visualisations & tidying up some of the titles, and you should end up with a dashboard looking something like the below – well done!





☒ Huzzah!

Correct